EXTRACTION AND UTILISATION OF VNTR ALLELES

Glossary of Terms and Abbreviations

adapter	nucleotide sequences, usually comprising annealed complementary oligonucleotides, ligated to DNA fragments that allow specific amplification and manipulation of those fragments
AFLP	amplified fragment length polymorphism
allele	one of several possible alternative sequence variations at any one locus
amplimer	the product, or pool of products, generated by amplification with the adapter primer and an 'internal primer'
DNA	deoxyribonucleic acid
DNA fingerprint	the display of a set of DNA fragments from a specific DNA sample
GMS	genomic mis-match scanning
individual	a member of any species subject to investigation
heteroduplex	a duplex of two alleles derived from different individuals, sets of individuals or populations
heterozygygous	alleles at the same locus of each of the paired chromosomes in a diploid cell being different
homoduplex	a duplex of alleles derived from the same individual, set of individuals or population
homozygous	alleles at the same locus of the paired chromosomes of a diploid cell being identical
locus	a specific position on a chromosome
mis-match	one or more bases in a duplex that fail to form stable

	hydrogen bonds with opposing bases
NASBA	nucleic acid sequence based amplification
PCR	polymerase chain reaction
RAPD	random-amplified DNA markers
RDA	representational difference analysis
RFLP	restriction fragment length polymorphism
trait	a distinguishing feature or characteristic manifesting itself physically, chemically or biologically
TRAIT	Total Representation of Alleles that are Informative for a Trait
VNTR	variable number tandem repeat, also referred to as simple sequence repeats (encompassing all repeats of two or more nucleotides that may be continuous or interrupted by short non-repetitive sequence, including minisatellites and microsatellites).

Field of the invention

The field of this invention is the detection of polymorphic variation in complex genomes, which is the mainstay of the study of hereditary traits in all organisms. Since polygenic traits far outweigh those that are monogenic, a procedure that allows the isolation in concert of several informative polymorphisms within the complex genomes of multiple individuals would provide an extremely powerful tool for the investigation of hereditary traits.

The invention differs fundamentally from all other techniques that have been previously employed by:

(i) permitting mass generation of VNTRs quickly and easily from DNA

(ii) generating polymorphisms that are both linked and informative for a trait;

(iii) reproducing and preserving the polymorphic allele, as it

occurs in the genome;

(iv) negating problems that are features of other polymerase chain reaction based techniques; including miss priming, reaction contamination and generation of spurious products;

(v) negating the need for investigations to be confined to families of closely-related individuals;

(vi) permitting the analysis of polygenic traits;

(vii) having a sparing requirement for DNA starting material.

The invention therefore represents a major advancement in the ability of workers in the biomedical fields to screen simple or complex genomes, rapidly and with fidelity, for polymorphisms co-segregating with advantageous or deleterious monogenic or polygenic hereditary traits. There is enormous potential for advancement of medicine, veterinary medicine, forensic science, agriculture, animal husbandry and biotechnology, by the generation of polymorphic markers co-segregating with hereditary disease or traits of social or economic importance. The invention will also serve to facilitate mutation analysis for all relevant organisms.

Introduction

DNA is a double stranded linear polymer composed of repetitions of four mononucleotide units. The sequence in which these units are arranged gives rise to a genetic code, referred to as the genome. Although the genomes of all individuals within a species are essentially homologous, subtle variations exist which impart individuality. Locations of the genome at which more than one sequence variation may exist are termed polymorphisms, each variant of that sequence representing an allele. Polymorphisms in gamete-forming germinal cells will be inherited by subsequent generations of progeny. By studying the combination of polymorphisms in the genome of an individual a unique code ('fingerprint') can be assigned and the ancestry of that individual can be determined.

Furthermore, a polymorphism found to be linked and co-segregating with a particular genetic trait or hereditary disease may be used as a marker for genetic screening of that trait or disease in other individuals.

The study of advantageous or deleterious hereditary traits in complex genomes has been the subject of considerable interest due to its economical, medical and social implications. The establishment of protocols that allow the comparison of nucleic acid sequences in complex genomes and the isolation of differences unique to a subset of those sequences is a fundamental requirement of this field of study.

A number of protocols have been used in animals and plants for the comparison of nucleic acid sequences and isolation of differences between those sequences in individuals. These protocols involve restriction fragment length polymorphism (RFLP), random-amplified polymorphic DNA markers (RAPD), amplified fragment length polymorphism (AFLP), representational difference analysis (RDA), genomic mis-match scanning (GMS), and linkage analysis of variable number tandem repeats (VNTR). These protocols detect polymorphisms by assaying subsets of the total DNA sequence variation in a genome. Polymorphisms detected by RFLP, AFLP, and RDA rely on the generation of a fingerprint ladder by gel-electrophoresis which reflects restriction fragment size variation. RAPD polymorphisms result from sequence variation at primer binding sites and differences in length between primer binding sites. GMS polymorphisms result from sequence variation within heterohybrid molecules comprising restriction fragments derived from two related individuals. Linkage analysis involves the detection of length variation of variable number tandem repeats (VNTRs) and co-segregation of one allele with a trait of interest.

RFLP

RFLP analysis relies on the cleavage of a nucleic acid sequence by restriction endonucleases and separation of the resulting

fragments by gel electrophoresis. The fragments are blotted onto a membrane and hybridized to labelled probes to allow detection of fragment length variation. This technique may be of use in the study of a single isolated locus or gene fragment, but where an investigation is not confined to an isolated sequence it is inadequate. Further limitations are that only a small number of the polymorphisms generated may be informative, there is a high demand for DNA starting material, and the method is labour intensive.

RAPD

RAPD is a commonly used PCR-based polymorphic marker technique in genomic fingerprinting and diversity studies, particularly for plant species. This technique involves the use of a single 'arbitrary primer' which gives rise to amplification of regions of genome where there is sufficient homology between the sequences of genomic DNA, in the 5' to 3' direction, and that of the arbitrary primer. The amplified products are separated by gel electrophoresis. Subtle variations of this method include arbitrary primed-PCR (AP-PCR) and DNA amplification fingerprinting (DAF). However, the principle of arbitrary priming and amplification of DNA by PCR for difference analysis is common to all. Advantages compared to RFLP are that these methods are more rapid, have a lower demand for DNA, and do not require prior knowledge of sequence. A limitation in common with RFLP is that each analysis can only compare the genomes of two individuals. Although several loci can be evaluated concomitantly by this method, detection of polymorphisms requires observation of variation in band patterns by gel-electrophoresis and is subject to errors of superimposition of different alleles of similar electrophoretic mobility. Many bands may be faint and difficult to interpret, and it is difficult to achieve consistent results in repeat experiments. In common with the majority of PCR techniques, the results are prone to error by subtle changes in reaction conditions, reagent contamination, and the generation of

inconsistent banding patterns. This lack of reliability limits the usefulness of such techniques in the 'typing' of individuals.

AFLP

AFLP analysis (EP, A, 0534858; Zabeau M *et al.*) involves restriction endonuclease digestion of DNA and ligation of the generated restriction fragments to adapters. Using primers complementary to the adapter sequence, the restriction fragments are amplified by PCR, and the products are separated by gel-electrophoresis, differences in band patterns revealing polymorphisms. Microsatellite-AFLP (WO 96/22388; Kuiper M *et al.*) is a modification of this technique in which two or more restriction enzymes, at least one of which cuts at a simple sequence repeat, are used to cleave DNA into fragments that are ligated to adapters. The fragments are amplified with primers complementary to the adapter sequence. In common with RAPD, several loci can be evaluated concomitantly by this method, but detection of polymorphisms requires observation of variation in band patterns by gel-electrophoresis and is subject to errors of superimposition of different alleles of similar electrophoretic mobility. The ability to score bands on an AFLP fingerprint is compromised by generation of large numbers of bands of which some may be very faint and difficult to interpret. Furthermore, the technique is prone to errors that are common to all PCR based techniques, summarised above, and suffers from an inability to analyse multiple complex genomes simultaneously. This is compounded by the generation of bands, by incomplete restriction of the template DNA, that do not reflect true polymorphisms. AFLP and RAPD analyses therefore share many of the same limitations. An additional problem is that AFLPs, rather than being evenly dispersed through out the genome, are reported to be clustered around centromeres. Consequently, this method may not allow the generation of polymorphisms that co-segregate with sequence differences of interest if they are located at a distance from centromeres. This problem is reflected in the reduced rate of

polymorphism detection compared to techniques such as linkage analysis. Furthermore, the complexity of the experimental data derived by AFLP becomes exaggerated with increasing complexity of the genome subject to analysis. Consequently, although it has been possible to investigate the genomes of some plant species by AFLP analysis, the relatively complex genomes of higher eukaryotic species may be beyond the useful capacity of this technique.

RDA

RDA involves restriction endonuclease digestion of DNA, ligation of the fragments to adapters and amplification by PCR. Differences between compared genomes are selected by successive rounds of subtractive hybridization and kinetic enrichment such that regions of difference predominate. This technique is prone to erroneous results through reaction contamination and generation of spurious products. In addition, a fundamental requirement of RDA is the availability of families of closely related individuals, some of which are manifesting the trait of interest. Where RDA is performed on anything other than closely related or highly inbred genomes the multiplicity of differences is too vast for succinct and useful analysis.

GMS

GMS is technique for mapping regions of identity-by-descent of two related individuals. The entire genome is compared in a single hybridisation that has a high demand for DNA since the genomic samples are not amplified. Freedom from the need of prior map information, conventional markers, or gel electrophoresis are to its advantage. However, the method is restricted to use on the genomes of only two related individuals.

Restriction fragments of the two genomes are hybridised, one of which having been methylated such that heterohybrid molecules can be

distinguished through their resistance to digestion by Dpn I and Mbo I that cleave only fully methylated and unmethylated molecules, respectively. Heterohybrids containing homologous strands that lack mis-matches are selected and used to probe an array of mapped clones. Although the mis-match proteins used in this technique may resolve point mutations polymorphisms comprising more substantial mis-matches that are beyond the limit of this system are not detected. Therefore, in keeping with RFLP, AFLP, RAPD, and RDA, GMS tends to resolve binary polymorphisms that may have low informative power.

In all of the above techniques it is essential that there is a difference in nucleotide sequence at or between primer binding sites or endonuclease restriction sites in order to detect polymorphisms. This highlights the major limitations of these procedures, because in many instances a mutation giving rise to a hereditary trait will not create a sequence difference detectable by variation in primer binding or restriction enzyme digestion. Consequently, a polymorphism linked to a trait of interest will not be identified using these techniques. GMS detects polymorphisms that are incidental to the restriction site and is spared some of the limitations of the other methods. However, in contrast to VNTR polymorphisms, the majority of polymorphisms detected by all of these techniques are not informative.

Linkage analysis

Linkage analysis is an indirect molecular genetic strategy that involves the systematic comparison of the inheritance of polymorphic VNTRs with the trait of interest in families in which that trait is present. There are a number of types of VNTR, including minisatellites and microsatellites, a feature of all being the repetition of elements of simple sequences. They are polymorphic by virtue of variation in the number of times each element is repeated, giving rise to alleles with variation in length. Since several alternative alleles may exist at any one locus, in

contrast to polymorphisms based on variation in primer binding or restriction enzyme digestion, VNTR polymorphic alleles tend to be highly informative. Consequently, where co-segregation of a trait with a particular VNTR allele is demonstrated, the allele may be used as a marker for that trait, or may be used as a vehicle to facilitate identification of the molecular genetic basis of the trait. Microsatellites are ubiquitously distributed throughout all eukaryotic genomes. Consequently, linkage analysis with microsatellites is associated with the highest polymorphism detection rate of the genetic screening methods. Indeed, systematic microsatellite analyses have already been responsible for many advances in the understanding of certain types of common cancer. Linkage analysis therefore has advantages compared to other related methods of difference analysis, the results of which are very reproducible. However, linkage analysis is very time consuming, labour intensive and expensive. Furthermore, since many analyses are performed individually the overall requirement for DNA is extremely high. This is particularly true if a physical map of the genome is unavailable for the selection of informative microsatellites that are evenly distributed throughout the genome. The demonstration of linkage requires the application of elaborate statistical programs and powerful computer software for analysis of the experimental data. This technique is better suited to monogenic defects since the statistical analyses required for multigenic traits are particularly complex. Unfortunately, multifactorial genetic traits are far more prevalent than monogenic defects, making linkage analysis a cumbersome technique for the investigation of the majority of hereditary traits.

The characteristics of an ideal protocol for isolation of polymorphisms co-segregating with disease in complex genomes would include:

(i) the ability to isolate simultaneously and with fidelity the polymorphisms from complex genomes of several individuals

(ii) the ability to isolate several polymorphisms simultaneously,

permitting the analysis of polygenic traits

(iii) a high detection rate of polymorphisms that co-segregate with sequence differences in all eukaryotic species, including subtle differences such as those resulting from point mutations

(iv) no requirement for large families of closely related individuals to study traits of interest

(v) no requirement for physical maps of the genome or prior knowledge of genomic sequence

(vi) a requirement for sparing quantities of nucleic acid samples for analysis

(vii) simplicity of use without a need for expensive specialist laboratory equipment or computer software

(viii) potential for widespread application throughout the animal and plant kingdoms

(ix) a robust performance with precision, accuracy and fidelity.

None of the techniques that are currently available fulfil the majority of these ideal characteristics. All are compromised by at least one of several limitations including: expense; lack of speed; requirement for large amounts of DNA; low polymorphism detection rate; an inability to detect small sequence variations such as point mutations; a lack of fidelity with high incidence of artefacts and spurious results; inability to analyse several complex genomes concomitantly; an inability to resolve simultaneously polymorphisms at multiple loci; an intrinsic need for closely related genomes for analysis; a need for prior knowledge of sequence; and complexity of analysis with a need for expensive equipment and computer software. In addition, those techniques that are reliant on large families of closely related individuals are further compromised where there are discrepancies in lineage, so that paternity testing may be an essential preliminary investigation to establish the integrity of each family individual subject to analysis.

The Invention

The invention is a novel method for generating *en masse* the VNTRs from genomic or synthetic DNA, while preserving each allele with its flanking sequence. These alleles may be used to produce a 'fingerprint' by gel electrophoresis, or they may be used as the starting material in protocols for genotyping individuals or protocols for isolation of polymorphic markers that co-segregate with hereditary traits. The latter may be achieved by mis-match discrimination to yield a pool of alleles that are common to all individuals manifesting a particular trait. Further mis-match discrimination of these selected alleles with the alleles of individuals in which the trait is not present, in solution or fixed to an array, allows purification of VNTRs with alleles that are both linked and informative for the particular trait. The end products, therefore, are designated a Total Representation of Alleles Informative for a Trait (TRAIT).

In one aspect the invention provides a method of making a mixture of VNTR alleles and their flanking regions of the genomic DNA of one or more members of a species of interest, which method comprises the steps of:

a) dividing genomic DNA of the species of interest into fragments,

b) ligating to each end of each fragment an adapter thereby forming a mixture of adapter-terminated fragments in which each 3'-end is blocked to prevent enzymatic chain extension,

c) using a portion of the mixture of adapter-terminated fragments as templates with an adapter primer and a VNTR primer to create a mixture of 5'-flanking VNTR amplimers,

d) using a portion of the mixture of adapter-terminated fragments as templates with an adapter primer and a VNTR antisense primer to create a mixture of 3'-flanking VNTR amplimers,

e) and using genomic DNA of the one or more members of the species of interest as template with the mixture of 5'-flanking VNTR

amplimers and the mixture of 3'-flanking VNTR amplimers as primers to make the desired mixture of VNTR alleles and their flanking regions.

The species of interest may be any eukaryotic species from the plant and animal kingdoms. Although they do not show repetitive sequences in quite the same way, prokaryotic species are also envisaged. An individual member of a species may be for example a plant or a micro-organism or an animal such as a mammal.

In another aspect the invention provides a portion of genomic DNA of one or more members of a species of interest, said portion consisting essentially of a representative mixture of alleles of a chosen VNTR sequence and their flanking regions.

The term "representative mixture of alleles" does not necessarily imply that all of the possible alleles, or even most of these possible alleles, of a chosen VNTR sequence are present. Whether a particular allele is present or not, e.g. in the mixture generated by the method defined above, may depend on the nature of a restriction enzyme used in step a) and on other factors.

The invention also provides a portion of genomic DNA of a species of interest, said portion consisting essentially of a representative mixture of 3'-flanking regions of a chosen VNTR sequence, each member of the mixture carrying an adapter at its 3'-end, and a representative mixture of 5'-flanking regions of a chosen VNTR sequence, each member of the mixture carrying an adapter at its 5'-end.

The invention also provides a method of treating nucleic acids which consist essentially of a mixture of polymorphic alleles, e.g. of a chosen VNTR sequence and their flanking regions, or alternatively a mixture generated in some other way such as AFLP, microsatellite-AFLP, GMS or RAPD, the mixture being representative of those which manifest a trait of interest, which method comprises separating and then re-annealing strands of the mixture, and

separating and discarding any mis-matches. Preferably the method comprises the additional step of hybridizing the said mixture with a mixture of corresponding polymorphic alleles, e.g. of the chosen VNTR sequence and their flanking regions, or alternatively a mixture generated in some other way such as AFLP, microsatellite-AFLP, GMS or RAPD, which are representative of those which do not show the trait of interest, and selecting mis-matches to provide a mixture of polymorphic alleles which are characteristic of the trait of interest.

The invention also provides kits comprising protocols and reagents for performing the methods herein described.

The salient points of the invention may be represented as follows:

(i) reduction in the complexity of the genome by double positive selection of genomic DNA restriction fragments that both ligate to a chosen adapter and contain a sequence with homology to a chosen primer, employing enrichment of such products by PCR, NASBA or other methods;

(ii) introduction of the selected enriched fragments to a genomic template in such a way that allows recreation of the VNTRs with the flanking sequences within that template, whilst preserving the allele and therefore the informativeness of each locus;

(iii) mis-match discrimination of the generated VNTR alleles to remove any spurious products of amplification that occur through miss priming events, reaction contamination, and subtle variation in reaction conditions;

(iv) selection of only those synthesised VNTRs alleles that are common to all individuals manifesting a particular trait or those alleles that predominate in such a group of individuals. This is achieved by strand dissociation and hybridization, giving rise to mis-match containing heteroduplexes of alleles at any locus that differ among the individuals. These complexes can be rejected by mis-match discrimination. The enriched alleles that are common to individuals manifesting the trait or

predominate in that group are sufficiently pure to be used as starting material in other DNA based studies that utilise polymorphic alleles;

(v) rejection of those alleles common to all individuals manifesting a particular trait or predominating in such a group that are also common to individuals in which the trait is not present. This is achieved by strand dissociation and hybridization of the VNTR alleles that are common to individuals manifesting a particular trait of interest or predominating in that group with the VNTR alleles of individuals in which the trait is not present followed by a further round of mis-match discrimination. In this case mis-match containing heteroduplexes and homoduplexes derived from the individuals manifesting the hereditary trait are selected. These represent polymorphic VNTRs with an informative allele that co-segregates with the particular trait of interest. Amplification of these VNTRs from the DNAs of individuals manifesting the trait of interest yields the informative alleles that may be used as DNA markers.

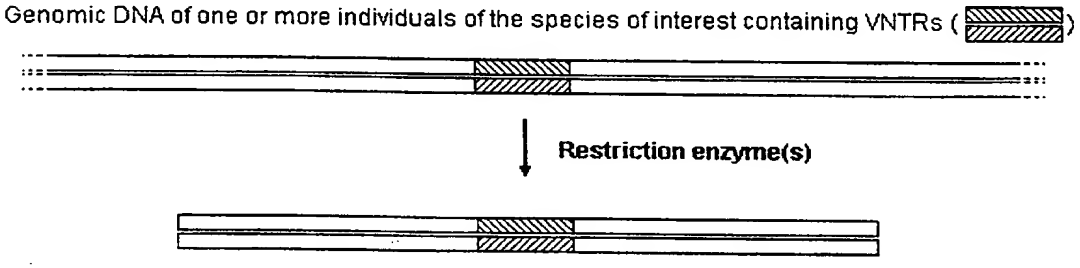
The invention provides a method of selecting genetic elements that are common to one pool of individuals but are absent in a second or present at a lower level. An obvious variation on this theme is the selection of genetic elements that are absent in one pool of individuals but are present in a second by judicious selection, during the course of the procedure, of allele duplexes that are either with or without a mis-match.

For simplicity, the protocol may be considered in three separate sections: generation of VNTR alleles; mis-match discrimination; and selection of alleles informative for a trait. The text is illustrated with a number of diagrams to facilitate description of the invention.

Generation of VNTR alleles

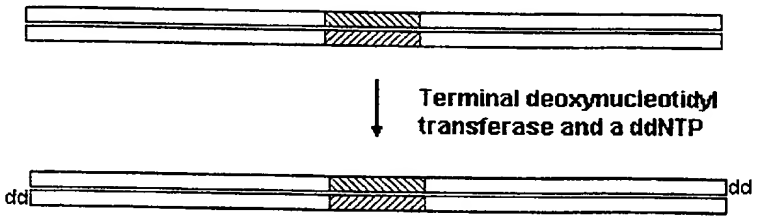
The protocol describes a method of generating with fidelity the VNTR alleles with their flanking sequences *en masse* from the genomic DNA of one individual, or the pooled DNAs of several individuals. The initial step involves fragmentation of the genomic DNA physically,

chemically or enzymatically, the aim of which is to obtain genomic fragments that contain VNTRs all of which being of an amplifiable length. The use of one or more restriction enzymes gives rise to uniform fragmentation of the genomic sample and constitutes the preferred technique. With judicious choice of restriction enzymes that cut frequently there is potential for generation *en masse* of every VNTR of the chosen type within a genome or pool of genomes since virtually all fragments will be sufficiently small for efficient amplification. It should be noted that the phenotype of individuals contributing genomic DNA for this fragmentation is unimportant. Indeed, the genomes restricted in this way need not be derived from any individual, or pool of individuals, that have been selected by virtue of their phenotype for investigation of a particular trait of interest.

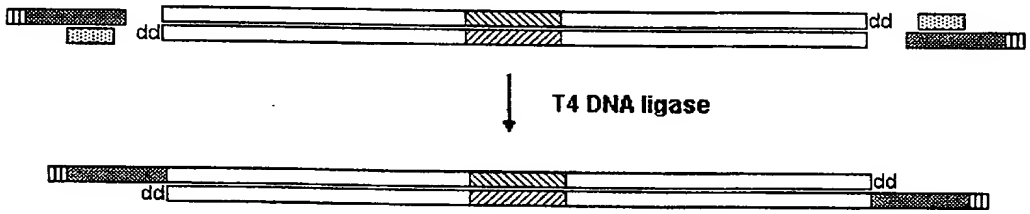


The restriction fragments are ligated to an adapter by which the fragments may be amplified or manipulated. The sequence of the longer oligonucleotide contained within the adapter is chosen such that it fails to generate any products when added as the primer to an amplification reaction containing genomic DNA as template. Termini are introduced physically, chemically or enzymatically to all available 3' ends to prevent their extension under the influence of a DNA polymerase. They may be introduced in one of several ways including: (A) addition of the terminus prior to ligation; (B) addition of the terminus following ligation; (C) addition of the terminus during ligation. The spectrum of available termini that are suitable for this purpose include, but are not limited to, dideoxynucleotide triphosphates.

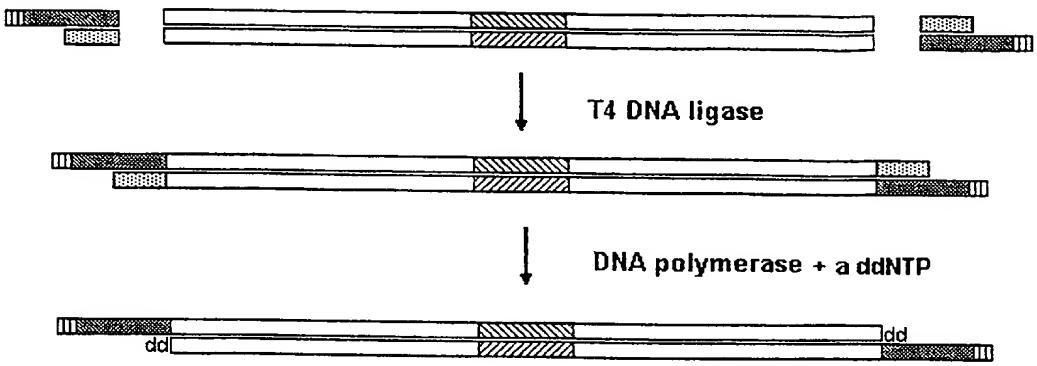
(A) A method by which termination may be achieved of all 3' ends with dideoxynucleotide triphosphates prior to ligation is through the action of a DNA polymerase, including Terminal deoxynucleotidyl transferase, in the presence of a chosen dideoxynucleotide triphosphate.



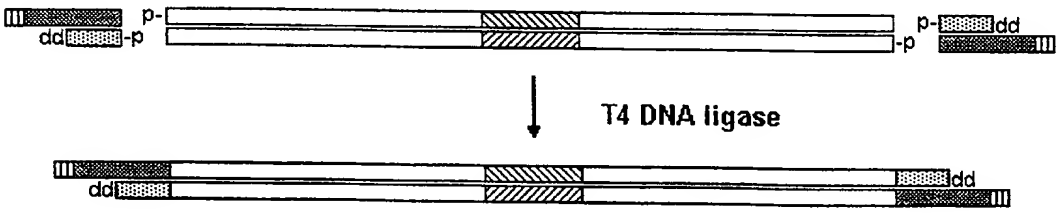
Ligation then follows with an adapter containing an appropriate 5' recess that accommodates the dideoxynucleotide triphosphate terminus on each strand.



(B) A method by which termination may be achieved of all 3' ends with dideoxynucleotide triphosphates following ligation is through the action of a DNA polymerase in the presence of a chosen dideoxynucleotide triphosphate.



(C) A method by which the ligated 3' ends can achieve termination during the ligation process is through incorporation of a suitable 3' terminus and a 5' phosphate on the shorter oligonucleotide during its synthesis such that this oligonucleotide will form a covalent bond with the genomic fragments under the influence of a enzyme such as T4 DNA ligase. Again, suitable termini include but are not limited to dideoxynucleotide phosphates, there being a variety of other modifications and deoxynucleotide analogues that will prevent extension of the 3' ends under the influence of a DNA polymerase.

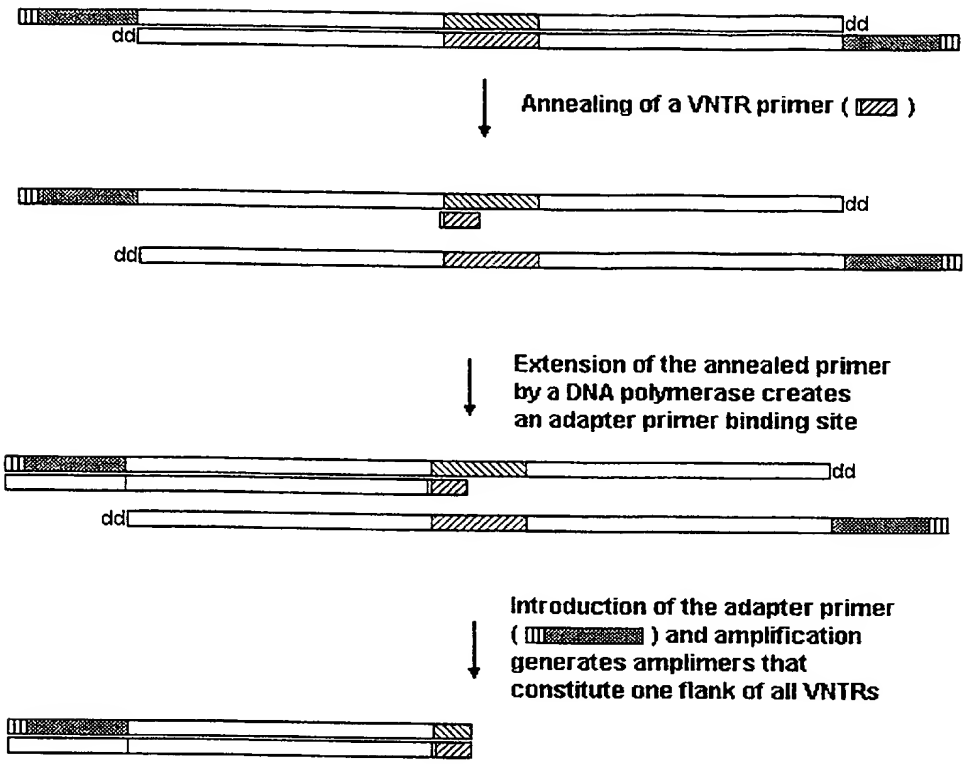


Of these, method (A) was found to be the most reliable since every genomic fragment that achieves ligation to an adapter is guaranteed to have an appropriate terminus. In addition, it guarantees that inter-fragment ligation is impossible. Method (C) also guarantees that each ligated 3' end possesses a terminus. However, unlike in the case of method (A), inter-fragment ligation can occur.

Since it is likely that some fragments will contain sites at which one DNA strand is nicked, in order to prevent polymerisation from these sites it is preferable to incorporate into them suitable termini. This may be achieved in a number of ways including, but not limited to, the incubation of the terminated and ligated genomic fragments with a DNA polymerase in the presence of all dideoxynucleotide triphosphates.

The longer oligonucleotide that is contained within the adapter may be used as the adapter primer in amplification reactions containing the genomic fragments that have been appropriately ligated and blocked by addition of termini at all potential sites of polymerisation. However, in the absence of 'internal' priming from another nucleotide sequence, the amplification of DNA is impossible. However, if another nucleotide sequence successfully anneals and achieves polymerisation to the limit of the adapter, an adapter primer binding site is created. Binding of the adapter primer will allow polymerisation of DNA to the limit of the annealed nucleotide sequence. If the nucleotide sequence represents a primer, or represents a nucleotide sequence containing a primer binding site, introduction of the adapter primer and the 'internal primer' allows specific exponential amplification of products only from those fragments that successfully ligated to the adapter and contain DNA homologous to that of the annealed nucleotide sequence.

If an oligonucleotide with sequence homology to a chosen VNTR is used as the internal primer, only those fragments that have ligated successfully to the adapter and contain the targeted VNTR will be capable of amplification. This gives rise to 'amplimers' that flank each VNTR, comprising genomic sequence limited by a restriction site for the chosen restriction enzyme and VNTR sequence with homology to the chosen VNTR primer.



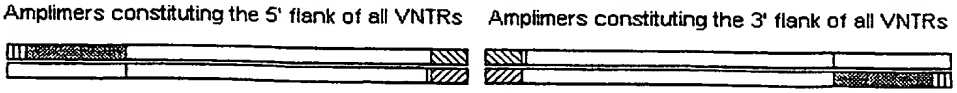
A number of different types of VNTR sequence have been identified in a diverse range of species. These include, amongst others, the dinucleotide repeats, trinucleotide repeats and the tetranucleotide repeats. Since the (AC)n dinucleotide repeat constitutes the most common VNTR that occurs in the majority of species, primers of appropriate sequence to generate amplimers for this VNTR may be chosen. It can be seen that the introduction of an (AC)n primer will give rise to amplimers that represent one flank of the VNTRs, and introduction of a (GT)n primer will give rise to amplimers that represent the other flank of these VNTRs. However, VNTRs with long repeat lengths will be over represented in the amplimer pool relative to shorter VNTRs by virtue of their greater number of primer binding sites. Similarly, the longer alleles will be over represented relative to the shorter alleles of the same VNTR due to their greater number of primer binding sites. This problem is negated by the introduction of

degenerate 3' ends on the VNTR primers that prevent polymerisation of the annealed primers unless they are aligned with the start of the flanking sequence. The amplification of all VNTRs and all alleles, therefore, will not be biased by their repeat lengths. In the case of (AC)n dinucleotide repeats the following primers may be used:

(AC)nB, where B = C + G + T
(CA)nD, where D = A + G + T
(GT)nH, where H = A + C + T
(TG)nV, where V = A + C + G

Alternatively, amplimers of other VNTR sequences may be generated in this manner by introduction of the appropriate target-specific primer containing a degenerate 3' end. Indeed, amplimers constituting genomic sequence that contain or flank any target-specific nucleotide binding site may be generated in the same way.

In the case of (AC)n dinucleotide repeats, the amplimers derived from reactions primed by the (AC)nB and (CA)nD degenerate oligonucleotides may be pooled. An obvious alternative is to generate an amplimer pool by priming amplification reactions with the (AC)nB and (CA)nD degenerate oligonucleotides together. However, this is likely to be less efficient than performing the reactions separately. Similarly, the (GT)nH and (TG)nV primed reactions may be pooled, or reactions containing both of these degenerate primers may be performed. Thus, two amplimer pools may be created, each representing sequences from only one flank of each VNTR.

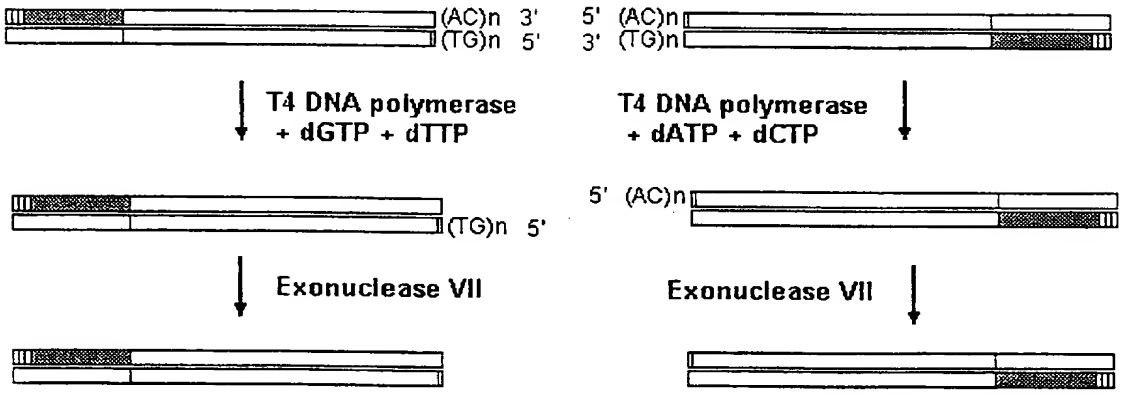


Since only one of the two flanking sequences of all VNTRs is generated in each amplimer pool, the full allele length being absent, the products of amplification are non-informative. However, the full length alleles, together with their flanking sequences, can be recreated with fidelity *en masse* from genomic DNA by hybridisation of the amplimers to that genomic DNA and subsequent polymerisation of the annealed sequences. As such, the full length 'affected' VNTR alleles of individuals manifesting a particular trait of interest may be obtained by hybridisation of the amplimers to the genomic DNAs of those individuals. Similarly, the reciprocal reaction for individuals in which that trait is absent will give rise to the generation of full length 'wild type' VNTR alleles and flanking sequences as they occur in the genomes of those individuals. Thus, two pools of VNTRs can be generated containing alleles derived from 'affected' DNA and alleles derived from 'wild type' DNA. A DNA polymerase that is highly processive is preferred in this application in order to minimise the potential for generation of 'stutter bands' that result from strand slippage during polymerisation.

To limit the potential for generation of spurious products by 'cross-talk' that occurs through the non-specific association of amplimer strands during hybridisation, it is preferable to remove the VNTR repeat sequences from the amplimers since these repeat sequences will be responsible for the majority of such cross-talk. This may be initiated in a number of ways including, but not limited to, (A) digestion by an enzyme with 3' to 5' exonuclease activity; (B) digestion by an enzyme with 5' to 3' exonuclease activity ; (C) digestion by Uracil DNA glycosylase of an amplimer pool generated with primers containing uracil; (D) digestion by RNase of an amplimer pool generated with an RNA primer.

(A) Providing the 5' end of the adapter primer has all four nucleotides represented the opposing strand will be similarly endowed. As such, incubation with an enzyme with 3' to 5' exonuclease activity, such as T4 DNA polymerase at 12°C in the presence of only two deoxynucleotide

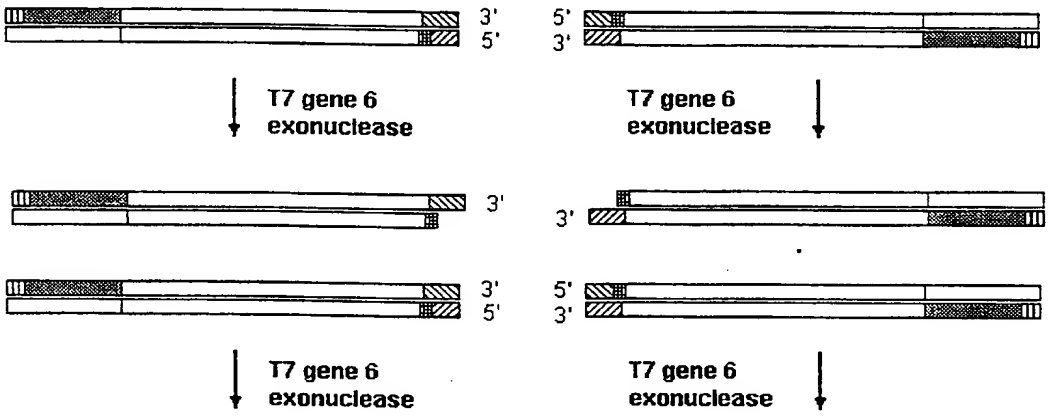
triphosphates, will not lead to significant shortening of the 3' strand complementing the adapter primer. The 3' strand complementing the VNTR primer, however, will be removed by T4 DNA polymerase if the reaction occurs in the presence of the deoxynucleotides that it lacks. Exonuclease digestion by the enzyme will cease when the first deoxynucleotide that is present in the reaction mixture is encountered. The 5' overhang that is created may be digested with a single strand specific exonuclease or endonuclease, including but not limited to Exonuclease VII, such that all repeat sequence is removed. The illustration depicts a scenario for (AC)n and (GT)n primed amplimers:



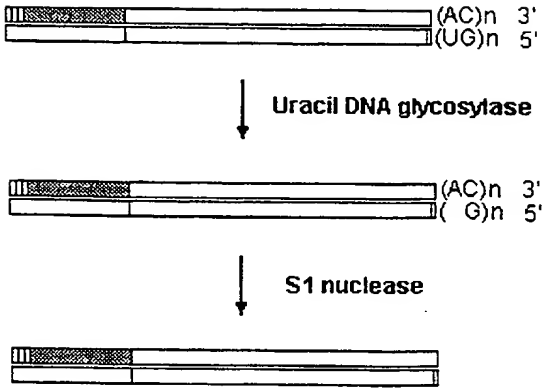
If a trinucleotide VNTR has been targeted appropriate digestion by T4 DNA polymerase in the presence of only one deoxynucleotide will be required. For tetranucleotide repeats this method is inappropriate and another should be adopted.

(B) The repeat sequence may be digested with a 5' to 3' exonuclease, such as T7 gene 6 exonuclease. Phosphorothioate bonds retard the activity of this enzyme. Four successive bonds are believed to inhibitory. Therefore, if the adapter primer has been synthesised with at least four phosphorothioate bonds at its 5' end, if not synthesised

completely with phosphorothioate bonds, it will be resistant to the 5' to 3' exonuclease activity of T7 gene 6 exonuclease. If the VNTR primers are synthesised with four phosphorothioate bonds at their 3' ends, the action of T7 gene 6 exonuclease will digest the VNTR primer leaving four nucleotides of repeat sequence. The complementary sequence may be digested by a single strand specific exonuclease or endonuclease, including but not limited to Exonuclease I, such that all repeat sequence is removed from the amplimers apart from four nucleotides in each strand. Such a short length of repeat sequence is unlikely to invite the generation of spurious products by non-specific interaction of strand ends during hybridisation.



(C) Synthesis of uracil containing VNTR primers, e.g. (GU)nH and (UG)nV, allows the destruction of these primers in the appropriate amplimer pool by the action of Uracil DNA glycosylase. Incubation of the digested amplimers with a single strand specific endonuclease, including but not limited to S1 nuclease, leads to further digestion of the VNTR primers that contains single stranded spaces and ultimately to the removal of the complementary sequence such that all repeat sequence is removed.

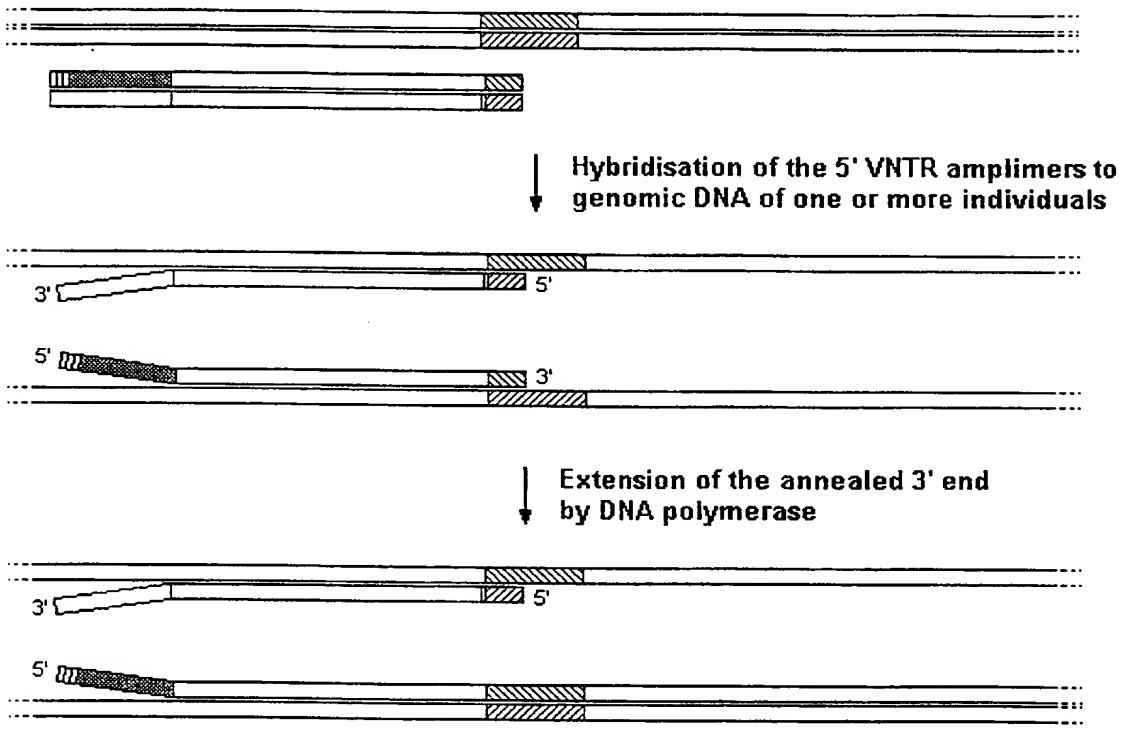


(D) The generation of amplimer pools with RNA primers based on VNTR sequence, using a DNA polymerase with reverse transcriptase activity, permits the destruction of the VNTR primers by the action of RNAse. The complementary sequence may be removed by a single strand specific exonuclease or endonuclease.

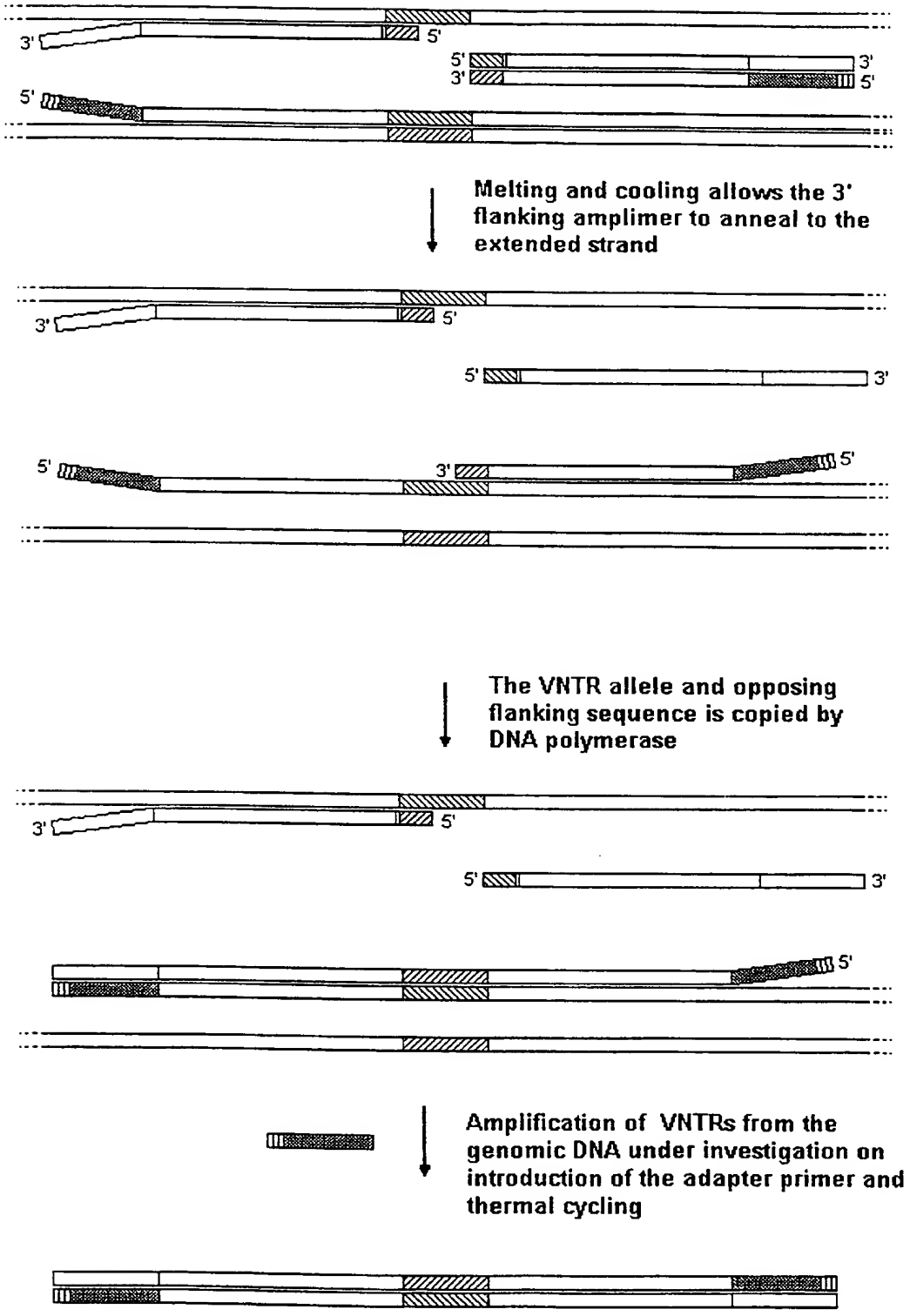
There are several methods by which the digested amplimers may be hybridised to the genomic DNA of one or more individuals to generate *en masse* and with fidelity the VNTR alleles as they occur in that template. These include (A) hybridisation and polymerisation of the amplimer pools, either separately in succession or together to genomic DNA that may or may not have been fragmented; (B) hybridisation and polymerisation of the amplimers constituting only one flank of each VNTR to genomic DNA that has been fragmented physically, chemically or enzymatically, and then terminated and ligated to an adapter which may or may not be the one used to generate the amplimer pools. In each case, the addition of one of many hybridisation accelerators will enhance the rate of hybridisation. Particularly under stringent conditions of hybridisation the use of such accelerators may be preferable. The number of methods by which hybridisation may be accelerated is vast but includes the incorporation of phenol exclusion, cationic detergents such as cetyl trimethylammonium bromide (CTAB), and volume excluding agents such

as dextran sulphate. It should be noted that if CTAB is the chosen hybridisation accelerator the salt concentrations in the hybridisation mixture should be low in order to prevent its precipitation.

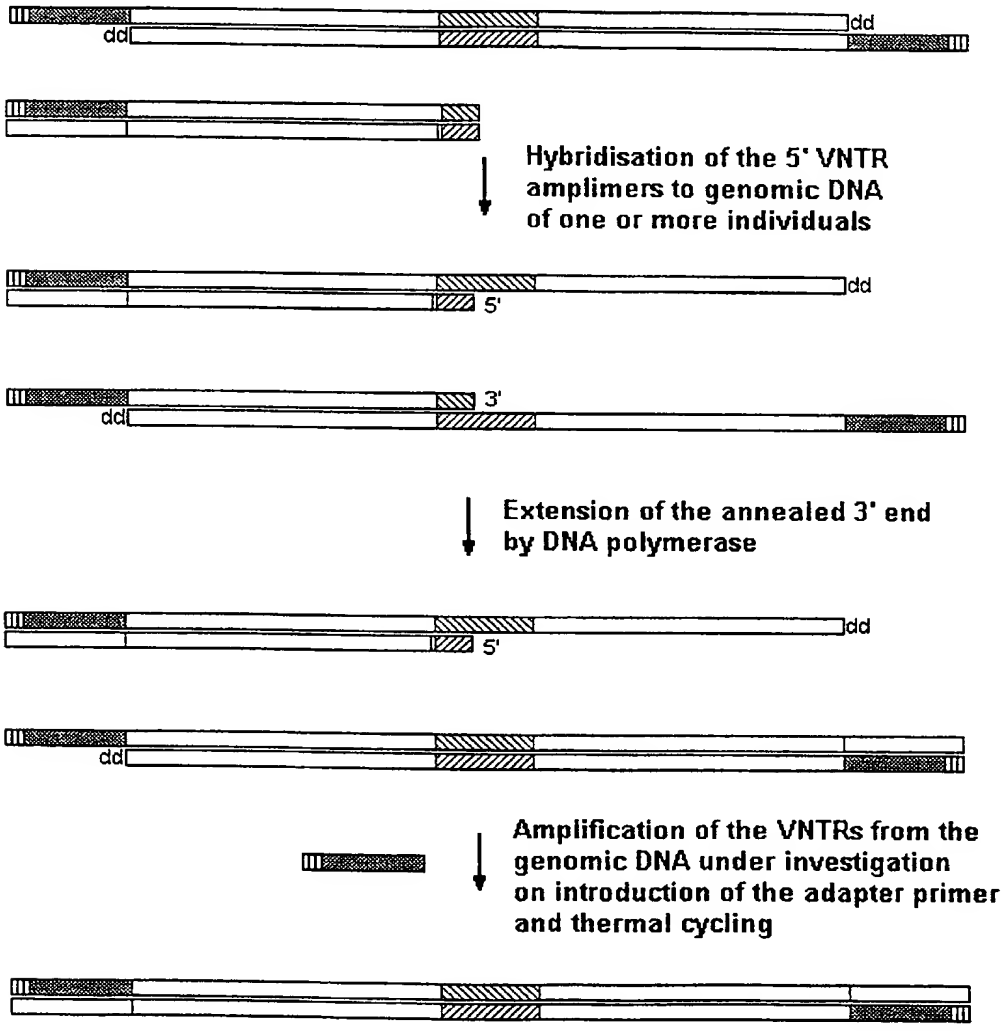
(A) Illustration is given for hybridisation of one amplimer pool to genomic DNA to permit the reproduction of VNTR alleles in that genomic template by a DNA polymerase:



Hybridisation of the second amplimer pool permits amplification of all VNTR alleles *en masse* using the adapter primer:


3'
5'
5'
3'
3'
5'
5'
Melting and cooling allows the 3' flanking amplimer to anneal to the extended strand
3'
5'
5'
3'
5'
3'
5'
The VNTR allele and opposing flanking sequence is copied by DNA polymerase
3'
5'
5'
3'
5'
Amplification of VNTRs from the genomic DNA under investigation on introduction of the adapter primer and thermal cycling

(B) Illustration is given for hybridisation of one amplimer pool to genomic DNA that has been fragmented, terminated and ligated to an adapter that may or may not be the same as that as that present in the amplimer pools:



Removal of repeat sequence from the amplimers permits concomitant hybridisation of both amplimer pools to genomic DNA while limiting the possibility for generation of spurious products through non-specific strand association. The generation of spurious products is reduced

further by hybridising the amplimers that constitute each flank separately in succession. This allows the introduction of further steps to control non-specific strand association including the removal of non-hybridised strands by incubation with a single strand specific exonuclease or endonuclease between hybridisations. In the preferred technique only one amplimer pool, comprising one flank of each VNTR, is hybridised to terminated and adapter-ligated genomic fragments. As such, this negates any possibility of non-specific association between amplimer strands of different pools. If each amplimer pool is hybridised and polymerised separately in this manner, the products that are generated in each reaction should be identical. Therefore, these products may be combined.

Hybridisation of the amplimers to the pooled genomes of several individuals allows the generation of the VNTR alleles that they contain. If this is performed on the pooled genomes of individuals manifesting a particular trait, and also on those of individuals lacking the trait, the 'affected' and 'wild type' alleles that are present in those pooled genomes can be synthesised.

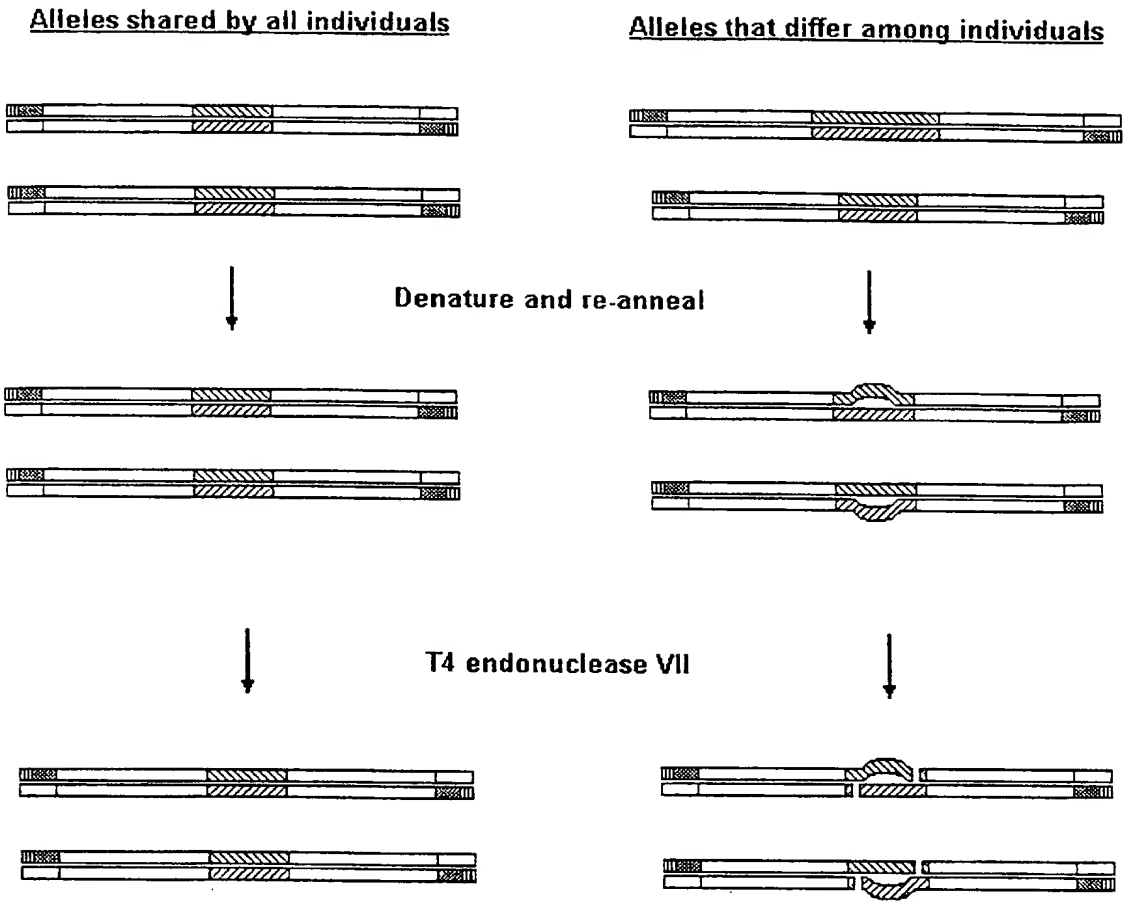
It is preferable to select the affected individuals from a defined population such that the same genotype is common to all individuals of a given phenotype. However, even if these individuals are selected from an out-bred population for which there are several genotypes that produce a single phenotype, the alleles that co-segregate with the trait loci will be present at a higher frequency in the pooled genomes of affected individuals than in the reciprocal pooled genomes of wild type individuals. These alleles will be enriched by successive repetitions of mis-match cleavage and amplification. To prevent the allele frequencies from being artificially skewed it is preferable to have a large number of individuals contributing genomic DNA to each pool. This ensures that the allele frequencies in the affected group and wild type group tend to equate to the general population from which they are derived such that disparity in the two is a consequence of linkage disequilibrium with the trait and not

another factor. However, if the numbers of affected and wild type individuals is limited the selection of matched sibling pairs, one member of each pair being affected and the other being a wild type individual, will go some distance to balance the allele frequencies of the pooled genomes other than with respect to the particular trait.

Mis-match discrimination

If the VNTR alleles that are generated from the affected individuals and the wild type individuals are denatured and allowed to re-anneal in separate reactions duplex DNA molecules with or without mis-matches will result. Due to the VNTR-specific flanking sequences and stringent conditions of hybridisation, only alleles that are of the same VNTR will re-anneal. Therefore, duplexes possessing mis-matches contain alleles of the same VNTR that are of unequal size or they contain spurious products of amplification. Alleles of similar size that re-anneal will form perfect duplexes.

The molecules that contain a mis-match may be digested with an enzyme that acts upon single stranded DNA or an enzyme that is able to detect conformational irregularities in DNA. Suitable enzymes include but are not limited to S1 nuclease and T4 endonuclease VII.

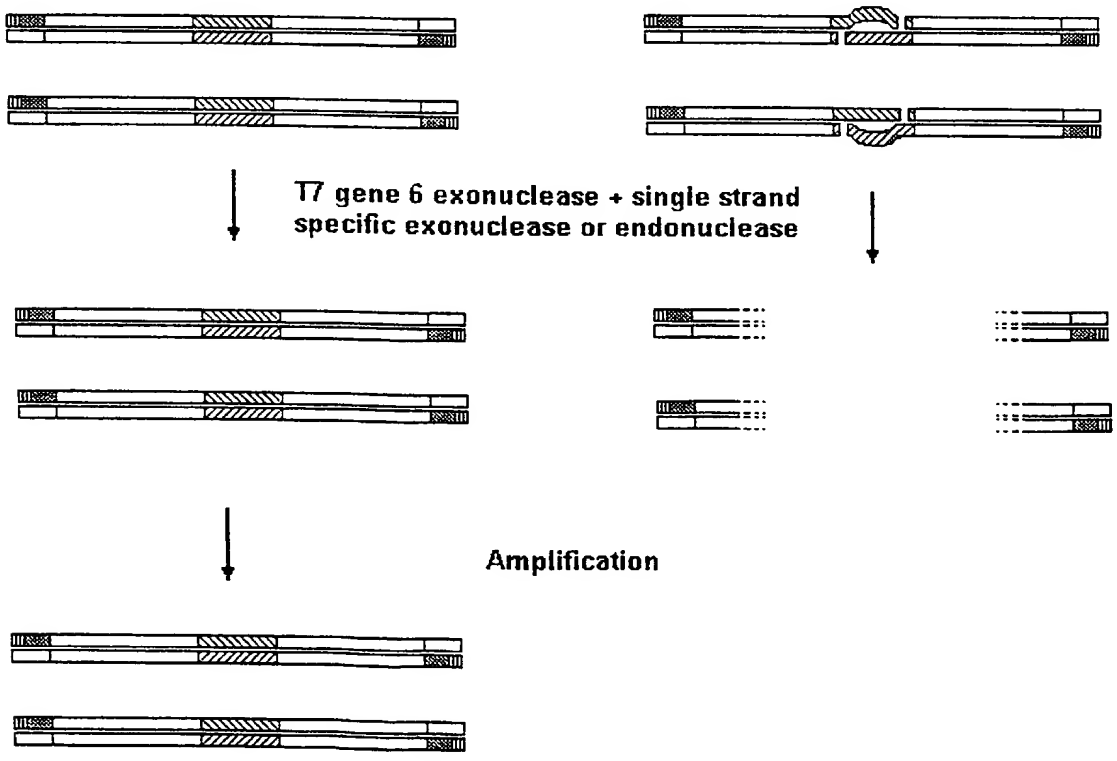


Of these two enzymes, T4 endonuclease VII has proved to be the most reliable and efficient enzyme in this application and has been found to digest efficiently in a range of DNA polymerase buffers while tolerating carry-over of CTAB from the hybridisation reaction. It cleaves both strands of a mis-match containing molecule leaving staggered ends, each strand being cleaved 3' with respect to the mis-match.

Cleavage is likely to occur within the repeat sequence creating ends that may interact non-specifically during the subsequent amplification process and resulting in the generation of spurious products. To obviate this problem the repeat sequences may be digested from the cleaved duplexes. This may be achieved in a number of ways, including

(A) by the action of a 3' to 5' exonuclease including but not limited to Exonuclease III, together with a single strand specific exonuclease or endonuclease, having protected all DNA strands prior to T4 endonuclease VII digestion with protective termini including but not limited to α-thiophosphate groups or a 3' overhang; (B) by the action of a 5' to 3' exonuclease including but not limited to T7 gene 6 exonuclease, together with an exonuclease or endonuclease, having protected all DNA strands prior to T4 endonuclease VII digestion with protective groups including but not limited to phosphorothioate bonds incorporated in to the adapter primer.

By inclusion of phosphorothioate bonds in the adapter primer the 5' ends of all molecules containing the adapter primer will be resistant to the 5' to 3' exonuclease activity of T7 gene 6 exonuclease. However, the 5' ends created by T4 endonuclease VII cleavage will be susceptible to this enzyme.



It is possible that some molecules will escape complete cleavage by T4 endonuclease VII acquiring merely a single stranded nick. However, such nicks are susceptible to digestion by T7 gene 6 exonuclease, though only the nicked strand would be digested if this enzyme was used in concert was a single strand specific exonuclease. On the other hand, a single strand specific endonuclease, including but not limited to S1 nuclease, would cleave the complementary single strand that is exposed by action of T7 gene 6 exonuclease in molecules receiving single stranded nicks such that both strands become disrupted. Thus, enzymes such as S1 nuclease in concert with T7 gene 6 exonuclease would lead to the complete digestion of all T4 endonuclease VII digested molecules irrespective of whether one or both strands was cut.

S1 nuclease has proven successful in this role, being capable of efficient digestion of single stranded DNA under alkaline conditions created by the T7 gene 6 exonuclease buffer. However, some non-specific digestion of DNA may occur with this enzyme. Since those molecules receiving single stranded nicks by the action of T4 endonuclease VII are likely to be few, it may be preferable to use a single strand specific exonuclease that is less likely to act in this way. Among such enzymes are included Exonuclease I and Exonuclease VII. Molecules that lack a mis-match are resistant to this regime of digestion and may be enriched by amplification. In order to minimise the generation of 'stutter bands' that result from strand slippage and polymerase errors during the amplification reaction, the number of cycles of amplification should not exceed that which gives adequate yields of product.

In addition to T7 gene 6 exonuclease, Exonuclease III may act at nicks in DNA molecules. In the absence of phosphorothioate bonds within the adapter primer this enzyme would create long 3' overhangs in nicked molecules on digestion to completion. Therefore, inclusion of a single strand specific endonuclease or exonuclease that would remove these overhangs would allow the elimination of the cleaved molecule

irrespective of whether T4 endonuclease VII disrupted one or both strands in a mis-match containing duplex. However, in order to obviate the need for the additional step comprising protection of the 3' ends of all DNA molecules prior to mis-match cleavage the use of T7 gene 6 exonuclease is preferred since protection of the 5' ends that is required for use of this enzyme is easily achieved by incorporation of phosphorothioate bonds into the adapter primer.

Another method by which cleaved molecules could be removed is by addition of a hapten, including but not limited to biotin-16-dUTP, at the sites of cleavage followed by physical separation of the cleaved molecules by the affinity of the hapten to another chemical. This could be achieved by termination of the 3' ends of all molecules prior to the mis-match cleavage procedure such that they are inert in the presence of a DNA polymerase. Suitable termini include but are not limited to dideoxynucleotide triphosphates which may be incorporated by a DNA polymerase including but not limited to Terminal deoxynucleotidyl transferase. Subsequent incubation of the cleaved molecules with biotin-16-dUTP in the presence of a DNA polymerase, such as Terminal deoxynucleotidyl transferase, will give rise to biotinylation of only those molecules which lack terminated 3' ends. Separation of the biotinylated molecules through binding to streptavidin could then follow.

In a similar manner, since molecules cleaved by T4 endonuclease VII have a 3' overhang these molecules could be removed through capture by single stranded binding proteins or chemicals that possess an affinity for single stranded DNA. It is likely that the overhang created by T4 endonuclease VII will be too small for efficient selection of the cleaved molecules by this method. However, they could be lengthened specifically by incubation with a DNA polymerase, including but not limited to Terminal deoxynucleotidyl transferase in the presence of one or more deoxynucleotide triphosphates, having terminated all 3' ends of the DNA molecules prior to mis-match cleavage with suitable termini that render

them inert in the presence of a DNA polymerase.

Physical separation of DNA molecules is cumbersome and relatively inefficient compared to separation by enzymatic means. Furthermore, the removal of molecules that possess single stranded nicks is likely to be unsuccessful. For these reasons methods of enzymatic differentiation of DNA species is preferred.

Reiteration of several rounds of denaturation, hybridisation and mis-match cleavage successfully eliminates all spurious products of amplification. Furthermore, it reduces to homozygosity all VNTRs such that only the most common allele of each VNTR remains, or it tends to eliminate those VNTRs for which many alleles are present with equal frequency. Rapid transition from the temperature of denaturation to that of annealing is required to prevent preferential annealing of identical sized alleles. This is may occur if the transition from the denaturation temperature to the annealing temperature is protracted. A hybridisation accelerator may be included to enhance the efficiency of hybridisation. This process carried out in parallel for the 'affected' VNTR alleles as well as the 'wild type' VNTR alleles will tend to achieve identical reduction to homozygosity and the generation of balanced allele frequencies. However, for a number of VNTRs the allele frequencies in the affected and wild type groups at the end of the mis-match cleavage procedure will be significantly different. Providing that the trait of interest is the only feature distinguishing the two groups of individuals from which the VNTRs were derived alleles that are over represented in the affected group relative to the wild type group must co-segregate with that trait. These are markers of the trait and should be selected.

The effect of reiterated mis-match cleavage on the allele frequencies of a VNTR can be illustrated with a basic scenario ignoring the efficiency of digestion, the effects of polymerase errors and the second order kinetics of hybridisation. Consider a VNTR for which three alleles are present as follows:

STARTING SCENARIO

Alleles	A	B	C
Allele frequency	$^{2}/_{4}$	$^{1}/_{4}$	$^{1}/_{4}$
Ratio	2	1	1

If the alleles are denatured and allowed to re-anneal duplex molecules with or without a mismatch will result. The proportion of each allele that forms a perfect duplex will depend on its allele frequency. All mis-match containing molecules theoretically would be susceptible to digestion by T4 endonuclease VII and would be eliminated. Thus, after the first round of mis-match cleavage the amounts and ratios of each allele remaining would be:

Alleles	A	B	C
Amount remaining	$^{4}/_{16}$	$^{1}/_{16}$	$^{1}/_{16}$
Total remaining	$^{6}/_{16}$		
Ratio	4	1	1
Allele frequency	$^{4}/_{6}$	$^{1}/_{6}$	$^{1}/_{6}$

After a second round of mis-match cleavage the allele frequencies would change further:

Alleles	A	B	C
Amount remaining	$^{16}/_{36}$	$^{1}/_{36}$	$^{1}/_{36}$
Total remaining	$^{18}/_{36}$		
Ratio	16	1	1
Allele frequency	$^{16}/_{18}$	$^{1}/_{18}$	$^{1}/_{18}$

After the 3rd round the theoretical allele frequencies would be as follows:

Alleles	A	B	C
Amount remaining	$^{256}/_{324}$	$^{1}/_{324}$	$^{1}/_{324}$
Total remaining	$^{258}/_{324}$		
Ratio	256	1	1
Allele frequency	$^{256}/_{258}$	$^{1}/_{258}$	$^{1}/_{258}$

Therefore, after two rounds one allele would predominate markedly. After a further round this allele would be present virtually exclusively. The ratio of the total amount of this VNTR remaining, relative to a VNTR for which there was only one allele prior to mis-match cleavage, would be:

$$^{6}/_{16} \times {}^{18}/_{36} \times {}^{258}/_{324} \; : \; {}^{1}/_{1} \times {}^{1}/_{1} \times {}^{1}/_{1}$$
$$= \; {}^{43}/_{288} \; : \; 1$$

In the same way the most common allele of any VNTR will predominate after a sufficient number of rounds of mis-match cleavage. Four rounds may be sufficient to reduce the VNTRs to near homozygosity, but the efficiency of enzyme digestion, the generation of polymerase errors and the kinetics of hybridisation are factors that will influence this. Disparity in the allele frequencies of affected and wild type VNTRs will lead to enrichment of different alleles in each group if the imbalance is sufficiently large. Such alleles are informative for the trait of interest but must be selected from other enriched alleles that may be identical in both the affected and wild type groups if these predominate in the population in general irrespective of the trait.

Further examples of mis-match discrimination under different scenarios is given in the Appendix.

Selection of alleles informative for a trait

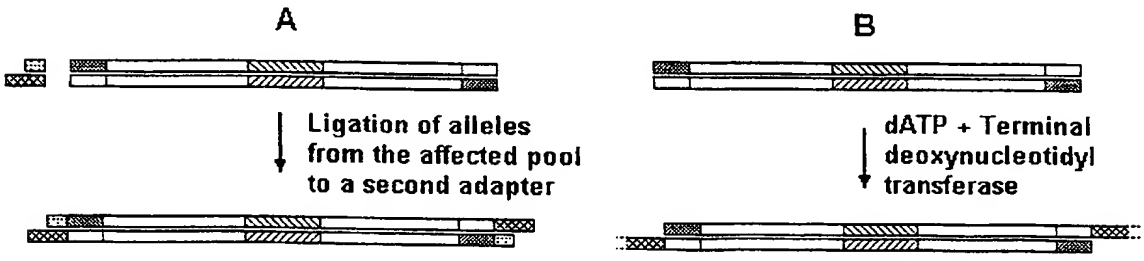
Selection of the alleles linked to the trait of interest may be achieved in a number of ways. Disparity in the allele size of each VNTR

surviving successive rounds of the mis-match cleavage procedure may be identified by hybridisation of these alleles from each group of individuals to an array of VNTR alleles of known length and spatial separation such that differences can be detected. Indeed, it may be possible to achieve quantitative hybridisation to an array in a similar manner that generates information regarding allele frequencies in the two groups without need of the mis-match cleavage procedure.

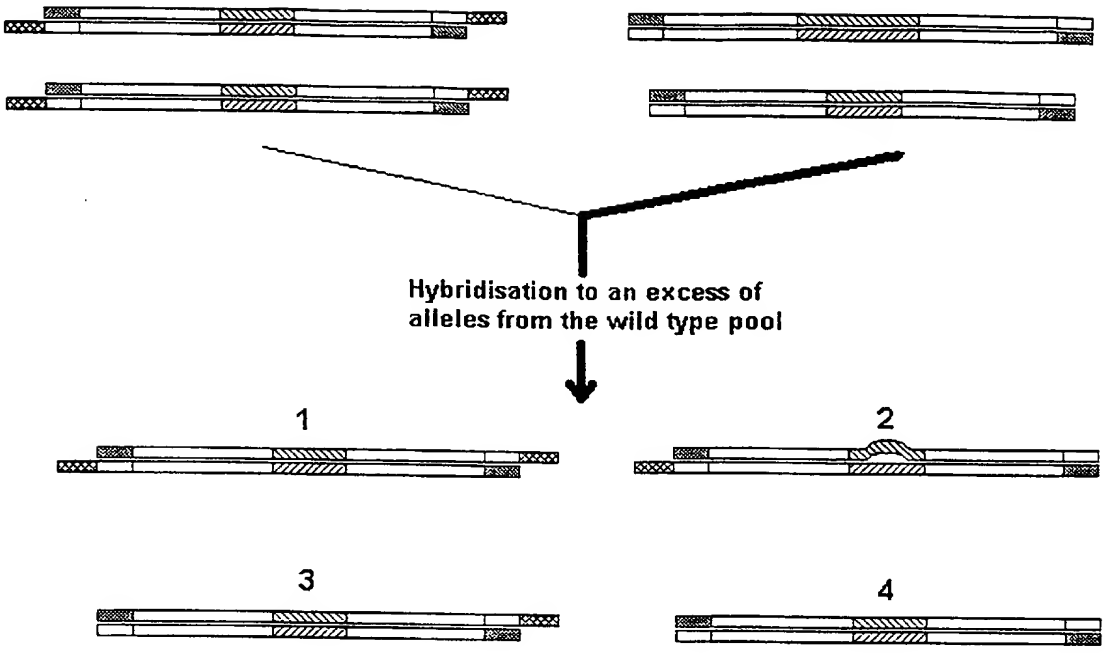
A less elaborate procedure involves the subtraction of the alleles in one group from those in another to identify differences in allele frequencies. However, this method must identify not only a VNTR for which an allele is present in one group but no alleles survive in the other group, but also a VNTR for which the alleles surviving in each group are different since both of these scenarios suggest linkage disequilibrium with the trait of interest. This can be achieved physically, chemically or enzymatically. If enzyme based selection is chosen it is preferable to amplify the alleles that have been enriched by the mis-match cleavage procedure with adapter primers that lack phosphorothioate bonds in order that enzyme digestion can proceed to completion.

A suitable method of enzyme based selection involves the addition of protective termini, including but not limited to a 3' overhang of at least four nucleotides or an α-thiophosphate linkage, to the surviving alleles of one group of individuals and subtraction with an excess of those surviving from the other group using Exonuclease III. Under most circumstances identification is required of any allele surviving from the affected individuals that fails to survive from those individuals lacking that trait. For this, addition of the protective termini should added only to the VNTRs derived from affected individuals. Obviously, the alternative strategy is possible. A 3' overhang may be created in a number of ways including but not limited to (A) ligation of an adapter, or by (B) non-template addition of nucleotides by a DNA polymerase. Of these, method (B) was found to be the more efficient which may be achieved using an enzyme

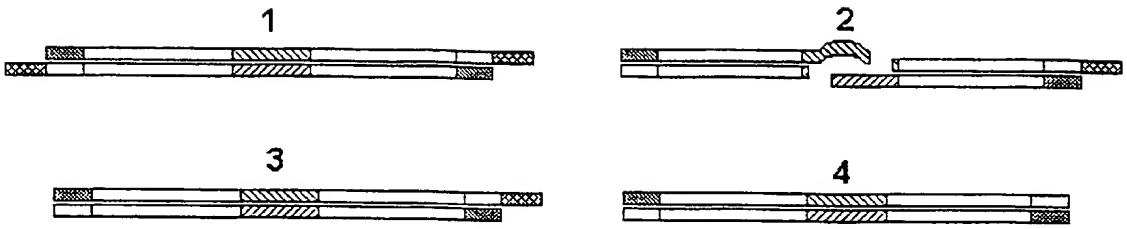
such as Terminal deoxynucleotidyl transferase. This enzyme may generate a 3' overhang of several hundred nucleotides on incubation in the presence of a single deoxynucleotide triphosphate. An α-thiophosphate linkage may be incorporated by addition of a protective deoxynucleotide analogue using a DNA polymerase including but not limited to Terminal deoxynucleotidyl transferase. Suitable analogues include α-thio deoxynucleotide triphosphates. Since these analogues may inhibit subsequent digestion or manipulation of the DNA molecules the addition of a 3' overhang to impart protection is preferred. Another less preferred method of imparting protection to the activity of Exonuclease III is through the action of an exonuclease with 5' to 3' activity, including but not limited to T7 gene 6 exonuclease, that may create a 5' recess in duplex DNA. The appropriate incorporation of phosphorothioate bonds within the adapter primer that is used to amplify the DNA molecules would ensure that digestion by T7 gene 6 exonuclease beyond that required to impart resistance to Exonuclease III is prevented. Similarly, a 5' recess could be created by incorporation of a uracil rich 5' end in the adapter primer which could be digested using an enzyme such Uracil DNA glycosylase.



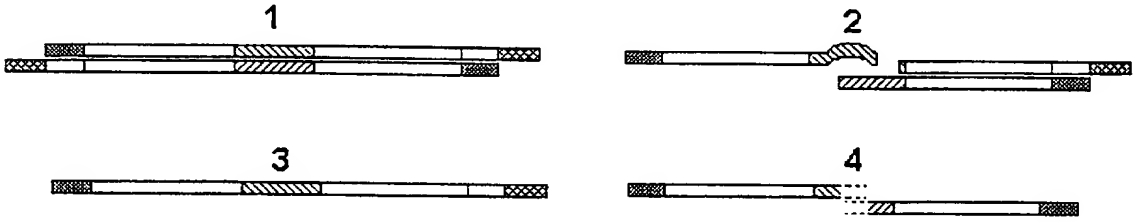
The resulting molecules are resistant to Exonuclease III digestion because of the 3' overhang that is created. Hybridisation to an excess of the surviving wild type alleles ensures heteroduplex formation of all affected alleles providing an allele of the appropriate VNTR survives in the wild type group.



If there are no wild type alleles to subtract from those of the affected group homoduplex molecules that possess a 3' overhang at each end will result (molecule 1). If the surviving allele of a VNTR differs between the two groups a heteroduplex molecule containing a mis-match will result (molecule 2). Surviving alleles of equal size in the two groups will give rise to heteroduplex molecules without a mis-match (molecule 3). The other species of DNA that will result from the hybridisation include homoduplexes of wild type alleles that may or may not contain a mis-match (molecule 4) and single stranded molecules that fail to hybridise. Digestion of these different types of molecule by an enzyme that acts on single stranded DNA or conformational irregularities in DNA, including but not limited to T4 endonuclease VII, results in cleavage of those duplexes containing a mis-match with the generation of a 3' overhang at the site of cleavage.



The subsequent digestion by Exonuclease III renders single stranded all duplexes or fragments of duplexes that do not possess a 3' overhang at each end.



Since the digestion of susceptible molecules by Exonuclease III tends to go to completion further digestion with a single strand specific exonuclease or endonuclease eliminates all single stranded DNA species and removes the 3' overhang on the surviving molecules. Therefore, only the target molecules survive digestion. Exonuclease I is suited to this task but often leaves a single nucleotide 3' overhang that must be removed if blunt end cloning is chosen as the means by which the target molecules are recovered.

For the intact homoduplexes the informative allele is present within the homoduplex and may be identified by cloning and sequencing. For T4 endonuclease VII cleaved fragments that have survived digestion by Exonuclease III and Exonuclease I, the full length VNTRs can be obtained by hybridisation of the fragments to fragmented, terminated adapter-ligated genomic DNA followed by amplification in a similar manner to that previously described. The informative allele may be identified by genotyping the individuals manifesting the trait of interest with respect to these VNTRs using VNTR-specific primers designed from their flanking sequences.

It is obvious that this method of subtraction is equally suited to other alleles besides those of VNTRs that may be generated in a variety of different ways. As such, this method of identifying differences in the composition of DNA pools may be applied more widely for selection of other types of polymorphic sequences as well as other species of DNA that may be present in one pool but absent in the same form in another.

This method is unique in its suitability for investigation of polygenic as well as monogenic hereditary traits. It is likely to make a significant impact in the study of hereditary traits, reducing considerably the difficulty, time and expense that is currently associated with this field of research.

The preferred embodiment

(i) Fragmentation of genomic DNA of an individual of the species under investigation, but not necessarily an individual in that investigation, with a single restriction enzyme.

(ii) Termination of all 3' ends by Terminal deoxynucleotidyl transferase in the presence of a dideoxynucleotide triphosphate.

(iii) Ligation of the terminated fragments to an adapter by incubation in the presence of T4 DNA ligase, followed by termination of single-stranded nicks.

(iv) Purification of the ligated products from the ddNTPs and amplification in reactions containing:

a) adapter primer and an (AC)nB primer, where B=G+T+C;

b) adapter primer and a (CA)nD primer, where D=G+A+T;

c) adapter primer and a (GT)nH primer, where H=A+T+C;

d) adapter primer and a (TG)nV primer, where V=G+A+C.

The products of amplification result from genomic fragments that successfully ligate to the chosen adapter and contain a VNTR with homology to the chosen primer.

(v) Digestion of the (AC)nB and (CA)nD primed products by T4 DNA polymerase in the presence of dATP and dCTP, followed by Exonuclease VII to remove all VNTR sequences and excess VNTR primer.

(vi) Digestion of the (GT)nH and (TG)nV primed products by T4 DNA polymerase in the presence of dGTP and dTTP, followed by Exonuclease VII to remove all VNTR sequences and excess VNTR primer. Size selection may be performed to obtain products of an optimal range of molecular weights.

(vii) Hybridization of an excess of either the combined (AC)nB and (CA)nD primed products or the combined (GT)nH and (TG)nV primed products with a sufficient amount of genomic DNAs derived from individuals manifesting a particular trait of interest.

(viii) Incubation of the hybridized products with Taq DNA polymerase to achieve strand extension of all annealed 3' ends.

(ix) Addition of adapter primer and generation of VNTR alleles from the 'genomic template' by thermal cycling in the presence of Taq DNA polymerase.

(x) Purification of the generated VNTR alleles followed by strand dissociation and reannealing under stringent conditions.

(xi) Digestion with T4 endonuclease VII of mis-match containing duplex molecules that result from hybridization of VNTR alleles to spurious products of amplification, or hybridization of VNTR alleles that differ among the individuals under investigation manifesting a particular trait of interest.

(xii) Further digestion by T7 gene 6 exonuclease together with S1 nuclease to remove VNTR sequence from cleaved molecules or eliminate them completely.

(xiii) Amplification of the surviving DNA molecules by thermal cycling in the presence of Taq DNA polymerase.

(xiv) Repetition of hybridization, digestion and amplification of the surviving DNA molecules. This enriches the reaction in VNTR alleles that are common to all individuals manifesting the particular trait of interest or those alleles that predominate in such a group and removes any spurious products of amplification.

(xv) Addition of a 3' overhang to the selected alleles of the group of individuals manifesting a particular trait by incubation with Terminal deoxynucleotidyl transferase in the presence of a dNTP.

(xvi) Hybridization of the selected VNTR alleles of the group of individuals manifesting a particular trait that possess a 3' overhang to an excess of the VNTR alleles of individuals in which the trait is absent that have been generated from their genomic DNAs in a method bearing similarity, wholly or in part, with (i) to(xiv).

(xvii) Digestion of mis-match containing duplex molecules by T4 endonuclease VII.

(xviii) Further digestion by Exonuclease III to eliminate strands in duplex molecules that lack protection by a 3' overhang.

(xix) Further digestion, after removal or inactivation of the Exonuclease III, by Exonuclease I to remove single stranded DNA. This results in elimination of all molecules other than the VNTRs linked to the

particular trait. For intact VNTRs the informative allele is present. For cleaved VNTRs that survive digestion by Exonuclease III and Exonuclease I the entire VNTR sequence may be obtained after hybridisation to fragmented, terminated, adapter-ligated genomic DNA and strand extension by Taq DNA polymerase such that VNTR specific primers may be designed from the flanking sequences that allow genotyping of affected individuals to implicate the informative allele linked to the trait.

A second embodiment

(i) VNTR alleles are generated by means other than processes of amplification of fragmented and ligated genomic DNA with adapter primer and VNTR primer, hybridization of the generated products to genomic 'template' DNAs of individuals manifesting a particular trait, and generation of the respective VNTR alleles from those template DNAs. These may include but are not limited to:

a) amplification of VNTRs from genomic or synthetic DNA using primers specific to the flanking regions of each VNTR in individual reactions;

b) amplification of VNTRs from genomic or synthetic DNA using a multiplex system, thereby allowing amplification of multiple VNTRs *en masse* using adapted VNTR specific primers;

c) amplification of VNTRs from genomic or synthetic DNA using an endonuclease that cleaves in or about VNTR sequences such that adapters may be ligated to the digested DNA and used for amplification of the VNTR alleles;

d) generation of a pool of VNTRs from individuals manifesting a particular trait by processes of subtraction with those in which the trait is absent.

(ii) Purification of the generated VNTR alleles followed by strand dissociation and reannealing under stringent conditions.

(iii) Digestion with T4 endonuclease VII of mis-match containing

duplexes that result from hybridization of VNTR alleles to spurious products of amplification, or hybridization of VNTR alleles that differ among the individuals under investigation manifesting a particular trait of interest.

(iv) Incubation of the hybridized alleles in the presence of T7 gene 6 exonuclease and S1 nuclease such that the digested duplex DNA molecules and single stranded DNA species are eliminated.

(v) Enrichment by amplification of mis-match free duplexes that are resistant to digestion.

(vi) Repetition of hybridization, digestion and selection of mis-match free molecules. This enriches the reaction in VNTR alleles that are common to all manifesting the particular trait of interest and removes any spurious products of amplification.

(vii) Hybridization of the selected VNTR alleles, that are common to all individuals manifesting a particular trait, to the VNTR alleles of individuals in which the trait is absent that have been generated from their genomic DNAs in a method bearing similarity, wholly or in part, with (i) to (vi).

(viii) Digestion with T4 endonuclease VII of mis-match containing duplexes followed by successive incubation with Exonuclease III and Exonuclease I.

(ix) Selection from the mixture of those surviving molecules that lack a 5' overhang. These entire VNTRs or VNTR fragments are linked to the particular trait of interest. The informative allele, with respect to the trait of interest, of the entire VNTRs can be established by sequencing. For the VNTR fragments the full length sequence can be generated by hybridisation to fragmented, terminated and adapter-ligated genomic DNA followed by incubation with Taq DNA polymerase. The informative allele may be established by various methods including but not limited to genotyping individuals manifesting the trait of interest using VNTR-specific primers designed from the flanking sequences.

Those that are skilled in the art will appreciate that there are

several methods of differentiating mis-match containing duplexes from those that are free of mis-matches, either in solution or on an array. The methods described in the above embodiments represent only one of these methods.

Those that are skilled in the art will appreciate that the invention is equally well suited any type of VNTR including but not restricted to dinucleotide repeats e.g.(CA)n and (GT)n, trinucleotide repeats e.g.(AAT)n, (AGC)n, (AGG)n, (CAC)n, (CCG)n and (CTT)n, and tetranucleotide repeats e.g.(CCTA)n, (CTGT)n, (CTTT)n.(TAGG)n, (TCTA)n, and (TTCC)n. In addition, the invention may be applied to simple organism microsatellites that include, but are not limited to, (AT), (CC), (CT) and (GA) rich tracts of repetitive motifs.

Those that are skilled in the art will appreciate that polymorphic alleles, other than those of VNTRs, may be used with the invention to produce alleles that are free of spurious products of amplification and are common to all individuals manifesting a particular trait. These polymorphic alleles may be hybridized to a fixed array of all possible alleles, or subset thereof, or to a pool of alleles derived from individuals in which that trait is absent. By mis-match discrimination those alleles linked and informative for a trait can be identified.

Those that are skilled in the art will appreciate that alleles from the genome of a single individual, or more than one individual, of unknown phenotype and genotype may be amplified with fidelity, removing the spurious products of amplification by mis-match discrimination, and hybridized to a fixed array of alleles, or to a pool of alleles in solution, in order assign a genotype or a phenotype to that individual.

Those that are skilled in the art will appreciate that mis-match discrimination may be performed using enzymes or chemicals other T4 endonuclease VII. These alternatives include but are not limited to S1 nuclease, Mung Bean nuclease, mutation detection proteins (e.g. Mut S), osmium tetroxide and hydroxylamine.

Those that are skilled in the art will appreciate that the polymorphic sequences that are amplified are themselves valuable and may be used in protocols other than that which determines co-segregation of VNTRs with a hereditary trait including but not limited to genotyping, mapping, positional cloning, quantification of trait loci, studies of ancestry and evolution, population studies, phylogenetics, and the study in vitro as well as in vivo of VNTRs and the sequences that separate them.

Those that are skilled in the art will appreciate that the invention may be used to identify somatic mutations that are non-hereditary if a VNTR is involved in that mutation.

Those that are skilled in the art will appreciate that the terminated and adapter-ligated genomic fragments may be used to recreate or amplify that region of the genome with sequence homology to any nucleotide sequence known or unknown to which they are hybridised.

Those that are skilled in the art will appreciate that the method represents a means of purifying a consensus sequence from PCR products such that the spurious products of amplification are eliminated.

Those that are skilled in the art will appreciate that the method represents a means of purifying a consensus sequence from any pool of one or more types of DNA molecule.

The invention differs fundamentally from all previous techniques since genomic fragments are generated that do not reflect the polymorphic variation at the locus from which they were derived. Furthermore, these fragments need not be generated from an individual in a particular investigation, but may be from any individual of the appropriate species. However, hybridization of these fragments to genomic 'template' DNA of an individual subject to investigation and mis-match discrimination permits amplification, with fidelity, of alleles within that genomic template whilst overcoming the problems of generation of spurious products that are a feature of other PCR-based methods. If the genomic fragments are derived from a single individual the problems of polymorphic variation

within the sequences that flank each VNTR are negated because these will be identical for all individuals under investigation. Since the invention preserves each VNTR allele with its flanking sequences, these alleles remain highly informative. In this respect the invention is unique. Furthermore, this novel method of generating VNTRs is rapid, inexpensive, has no requirement for prior knowledge of sequence, and has no requirement for elaborate equipment, it is of immense importance obviating the high investment of time and money that is currently required for isolation of VNTRs. Consequently, the application of technologies dependant on the availability of VNTR in species in which none have been isolated will be possible where previously this was unfeasible. The ability to generate large numbers of VNTRs from all species quickly, efficiently, cheaply and with fidelity is a considerable contribution of the present invention to workers in the to the biomedical field.

In summary, the invention involves a novel method of generating VNTRs encompassing restriction endonuclease digestion of DNA, ligation of the fragments to adapters and, by introduction of a primer with sequence homology to a chosen VNTR, amplifying only those fragments that are flanked by a chosen endonuclease restriction enzyme site and a VNTR. These fragments are not representative of the alleles of each VNTR and need not be generated from any specific individual under investigation. Hybridization of these fragments with genomic DNA of the individuals under investigation recreates the intact VNTR alleles with flanking sequence, as they occur in the genome. This in itself constitutes a major step in the ability of workers in the biomedical fields to generate quickly, efficiently, cheaply and with fidelity VNTRs in all species for purposes reliant on the availability of VNTRs, including but not confined to DNA fingerprinting and linkage analysis. The incorporation of a mis-match discrimination procedure overcomes the problems of miss-priming and generation of spurious products by reaction contamination and subtle variation in reaction conditions, that are to the detriment of all PCR-based

technologies, and allows exclusion of alleles that are not common to all individuals under investigation that manifest a particular trait. A second round of mis-match discrimination removes un-informative alleles that are present in the genomes of individuals that do not manifest the trait. This procedure is designated a Total Representation of Alleles that are Informative for a Trait (TRAIT). The invention, therefore, has significant advantages over previous methods, embracing the speed of analysis of AFLP, GMS, RDA and RAPD, and the high polymorphism detection rate of linkage analysis, but negating the need for DNA from closely related individuals and for paternity testing. The invention also overcomes fundamental problems that are a feature of PCR based technologies, including miss-priming and generation of spurious products through reaction contamination and subtle variations in the conditions of reaction. Furthermore, there is no requirement for expensive equipment or elaborate statistical computer software. The analysis will give rise to alleles that are both linked and informative, being present exclusively or at a higher frequency in individuals manifesting the trait of interest but absent or present at a lower frequency in those individuals that lack the trait. In this respect, the invention is unchallenged in its superiority over all other methods.

The invention allows concomitant detection of polymorphisms at multiple loci by simultaneous comparison of simple or complex genomes from multiple individuals and differs fundamentally from all other techniques that have been previously employed. The invention represents a major advance in the ability of workers in the biomedical fields to generate VNTRs from the genomes of any species quickly, efficiently, cheaply and with fidelity in addition to screening complex genomes for polymorphisms co-segregating with hereditary traits. Application of this procedure will therefore facilitate the development of markers for genetic screening for hereditary disease, or advantageous monogenic or polygenic traits in all organisms.

Examples of How the Invention may be Applied

The following illustrations represent examples of how the invention may be applied without inferring any limitation to scope of the invention or any limitation to the different ways in which the invention may be applied.

Experimental Data

Example 1

Preparation of amplimers using $(CA)_{13}$ and $(GU)_{13}$ primers.

2µg DNA was completely digested with 3µl *Rsa* I in a total volume of 100µl:

8.5µl	genomic DNA (equivalent to 3µg DNA)
10µl	10x reaction buffer
3µl	*Rsa* I (10u/µl; Promega)
78.5µl	dH_2O
100µl	

The reaction was incubated at 37°C over night followed by heat inactivated by incubation at 70°C for 20 minutes. The DNA was separated from the buffer by microconcentration (Microcon-100; Amicon). A volume of 10µl was recovered.

2nmoles of 48mer and 2nmoles of 12mer oligonucleotides that constitute the adaptor were combined:

15.9µl	48mer (equivalent to 2 nmoles)
13.7µl	12mer (equivalent to 2 nmoles)
10µl	10x ligase buffer (NEB)
48.4µl	dH_2O
88µl	

The mixture was heated to 50°C and allowed to cool to 10°C over 1 hour.

To the 88µl of annealed adaptor was added the 10µl of digested DNA and ligation of the adaptor to the genomic fragments was

performed:

88µl annealed adaptor/ ligase buffer (containing ATP)
10µl DNA
2µl T4 DNA ligase (400 NEBu/µl)
100µl

The reaction was incubated at 16°C over night and then heat inactivated by incubation at 70°C for 20 minutes.

The adaptor-ligated DNA fragments were separated from the buffer and non-ligated adaptor by microconcentration (Microcon-100; Amicon). A volume of 12µl DNA was recovered.

The adaptor-ligated DNA fragments were incubated with *Taq* DNA polymerase in the presence of dideoxynucleotide triphosphates to prevent 3' extension of the adaptor and non-ligated DNA in subsequent manipulations:

12µl microconcentrated DNA
3µl 10x NH4 reaction buffer
1µl 50mM $MgCl_2$
1µl 10mM ddATP
1µl 10mM ddCTP
1µl 10mM ddGTP
1µl 10mM ddTTP
1µl *Taq* DNA polymerase (5u/µl; Bioline)
9µl dH_2O
30µl

The reaction was incubated at 72°C for 2 hours.

The adaptor-ligated DNA with terminated 3' ends was purified by phenol/chloroform extraction and microconcentration. The volume recovered was made up to 40µl and the concentration of DNA was gauged by gel electrophoresis. A concentration of 75ng/µl was determined.

(CA) primed amplimers and (GU) primed amplimers were generated in separate reactions:

10µl 10x NH_4 reaction buffer

8µl 50mM $MgCl_2$

1.5µl 10mM dNTPs

1µl adaptor-ligated DNA with terminated 3' ends

4µl (CA) or (GU) primer (25pmol/µl)

73.5µl dH_2O

98µl

The reaction was overlaid with mineral oil and heated to 95°C for 2 minutes, during which time 1µl *Taq* DNA polymerase (5u/µl; Bioline) and 2µl adaptor primer (50pmol/µl) were added.

Thermal cycling was performed as follows: 95°C for 30 seconds, then 72°C for 45 seconds for a total of 20 cycles, followed by 72°C for 5 minutes.

To the 100µl of (CA) primed products was added 5µl Exonuclease I (10u/µl) to remove the remaining (CA) primer. This reaction was incubated at 37°C for 30 minutes.

To the 100µl of (GU) primed products was added 10µl Uracil-DNA glycosylase (1u/µl; NEB) to digest all uracil incorporated into the PCR products. This reaction was incubated at 37°C for 2 hours. 1µl 10mM dNTPs was added followed by 2µl T4 DNA polymerase (5u/µl; Epicentre laboratories) to remove the protruding (CA) strand that complemented the digested (GU) sequence. This reaction was incubated at 37°C for 5 minutes. Both the pools of amplimers were phenol/chloroform extracted and microconcentrated (Microcon-100; Amicon). For each pool, the volume recovered were made up to 500µl, of which 5µl was analysed by spectrophotometry to determine the concentration of DNA.

Equal amounts of (CA) and (GU) primed amplimers were hybridized to genomic 'template' DNA of a single individual prior to thermal cycling. In order to gauge the optimal ratio of amplimer to genomic 'template' DNA several reactions were performed using various amounts of

'template' DNA while keeping the amount of amplimers constant:

'Template' DNA (ng)	0	0.1	1	10	100	1000
Combined amplimers (ng)	1	1	1	1	1	1
5M NaCl (µl)	0.22	0.22	0.22	0.22	0.22	0.22
dH_2O (µl)	To a final volume of 5.55µl					

Each reaction was overlaid with mineral oil and incubated at 98°C for 5 minutes, after which the temperature was reduced stepwise to 78°C over 4 hours.

The following was added to each hybridization:

5µl 10X NH_4 reaction buffer

4µl 50mM $MgCl_2$

0.75µl 10mM dNTPs

0.5µl adaptor primer (50pmol/µl)

34.2µl dH_2O

Each reaction was spun briefly in a microfuge. They were heated to 72°C for 2 minutes and 0.5µl *Taq* DNA polymerase (5u/µl;Bioline) was added. The reactions were incubated at 72°C for a further 10 minutes, after which the temperature was raised to 95°C for 2 minutes. Thermal cycling was performed as follows: 95°C for 30 seconds, then 72°C for 1 minute, for a total of 10 cycles.

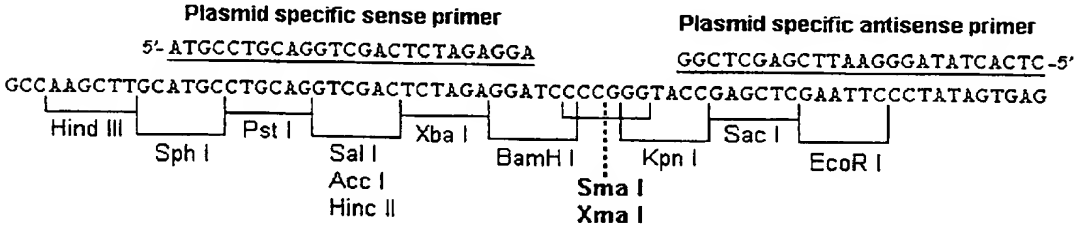
For each reaction 10µl of products amplified for 10 cycles were added to 40µl of reaction mix and amplified under the same conditions for an additional 22 cycles. 5µl of the ends products of amplification were run on an agarose gel. The reaction containing 100ng genomic 'template' DNA was found to yield the most products of amplification, equivalent to a ratio of 100:1 by mass of genomic 'template' DNA: amplimer.

The invention was validated by cloning the products of amplification. Two colonies of *E.coli* that had successfully transformed

were cultured, from which plasmids were later harvested. These plasmids were sequenced and were found to contain VNTR sequences at the multiple cloning sites.

Further experimental data

For the following experiments canine genomic DNA or cloned VNTR alleles amplified from canine genomic DNA were used. The cloned alleles were ligated into the Sma1 site of the pUC18 MCS, either side of which plasmid specific primers were designed for subsequent amplification of the plasmid inserts.



All reagents were obtained from Amersham Pharmacia Biotech, or its subsidiary companies, unless stated otherwise.

Oligonucleotides were obtained from Genset Corp., France. The VNTR primers (AC)11B, (CA)11D, (GT)11H and (TG)11V comprised eleven repetitions of the sequence shown in brackets followed by a degenerate base were B = C + G + T, D = A + G + T, H = A + C + T, and V = A + C + G.

Example 2

Generation of adapter-ligated, dideoxynucleotide terminated genomic fragments with (a) termination preceding adapter ligation and (b) adapter ligation preceding termination.

(a) 5μg canine genomic DNA were fragmented with Hae III, the digestion proceeding to completion over 12 hours at 37°C:

4.4µl 1.135µg/µl genomic DNA

10µl 10x restriction buffer

2µl 10u/µl Hae III

84µl dH_2O

100µl

Digestion was confirmed by electrophoresis of an aliquot of the reaction on a 1% agarose gel stained with ethidium bromide.

The DNA was extracted (GFX purification column) and eluted in 50µl 5mM Tris pH8.5, of which 30µl was incubated with Terminal deoxynucleotidyl transferase for 3 hours at 37°C :

30µl DNA

30µl 5x Terminal deoxynucleotidyl transferase buffer

4.5µl 10mM ddGTP

10µl 9u/µl Terminal deoxynucleotidyl transferase

75.5µl dH_2O

150µl

The DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. A volume of 35µl was recovered.

An adapter was prepared by annealing two oligonucleotides, a 24mer (GsCsAsGsGAGACATCGAAGGTATGAAC, SEQ ID NO: 4 where 's' represents a phosphorothioate bond) and a 12mer (TTCATACCTTCG; SEQ ID NO: 5):

7.6µl 197pmol/µl 24mer

9.2µl 162pmol/µl 12mer

1.87µl 10x T4 DNA ligase buffer

18.7µl

The mixture was heated to 55°C and allowed to cool to 10°C over one hour.

The adapter was ligated to the terminated genomic

fragments:

35µl DNA

18.7µl adapter

4.3µl 10x T4 DNA ligase buffer

1.5µl 10u/ul T4 DNA ligase

2.5µl dH_2O

62µl

The reaction was incubated at 16°C over night, then heat inactivated at 70° C for 20 minutes.

The DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. A volume of 54µl was recovered.

To prevent generation of spurious products through priming from sites of single strand nicks, these were terminated by incubation with Thermo Sequenase:

54µl DNA

4.4µl Thermo Sequenase buffer

1.4µl 10mM ddATP

1.4µl 10mM ddCTP

1.4µl 10mM ddGTP

1.4µl 10mM ddTTP

0.5µl 32u/µl Thermo Sequenase

5.5µl dH_2O

70µl

The mixture was overlaid with mineral oil and incubated at 74°C for 2 hours.

The DNA was extracted (GFX purification column) and eluted in 50µl 5mM Tris pH 8.5.

(b) 5µg canine genomic DNA were fragmented with Mbo I, the

digestion proceeding to completion at 37°C:

4.4µl 1.135µg/µl genomic DNA

10µl 10x restriction buffer

2.5µl 10u/µl Mbo I

83µl dH_2O

100µl

Digestion was confirmed by electrophoresis of an aliquot of the reaction on a 1% agarose gel stained with ethidium bromide.

Following incubation at 70°C for 20 minutes the DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. A volume of 32µl was recovered of which half was ligated to an adapter:

An adapter was prepared by annealing two oligonucleotides, a 24mer (GsCsAsGsGAGACATCGAAGGTATGAAC, SEQ ID NO: 4 where 's' represents a phosphorothioate bond) and a 16mer (GATCGTTCATACCTTC; SEQ ID NO: 6):

6.3µl 197pmol/µl 24mer

8.5µl 147pmol/µl 16mer

1.65µ 10x T4 DNA ligase buffer

16.5µl

The mixture was heated to 55°C and allowed to cool to 10°C over one hour.

The adapter was ligated to the genomic fragments:

16µl DNA

16.5µl adapter

2.4µl 10x T4 DNA ligase buffer

2µl 10ul/µl T4 DNA ligase

3.1µl dH_2O

40µl

The reaction was incubated at 16°C over night, then heat

inactivated at 70°C for 20 minutes.

The DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. A volume of 40µl was recovered.

The adapter-ligated fragments were terminated using Thermo Sequenase:

40µl DNA

4.4µl Thermo Sequenase buffer

1.4µl 10mM ddGTP

0.5µl 32u/ul Thermo Sequenase

24µl dH_2O

70µl

The reaction was overlaid with mineral oil and incubated at 74°C for 1 hour. To prevent generation of spurious products through priming from sites of single strand nicks, these were terminated by further incubation with Thermo Sequenase and addition of the remaining ddNTPs:

1.4µl 10mM ddATP

1.4µl 10mM ddCTP

1.4µl 10mM ddTTP

0.3µlThermo Sequenase buffer

4.8µl

The reaction was incubated at 74°C for a further hour.

The DNA was extracted (GFX purification column) and eluted in 50µl 5mM Tris pH 8.5.

Methods **(a)** and **(b)** of adapter ligation and termination of the genomic fragments were compared by amplification of the resulting fragments with or without an 'internal' primer in reactions comprising the following:

5µl	5µl	5µl	10x Taq PCR buffer
5µl	5µl	5µl	10x dNTPs
1µl	1µl	1µl	25pmol/µl 24mer
1µl	1µl	0µl	50pmol/µl (AC)11B
50ng	0ng	50ng	GFX extracted DNA
	to 50µl		dH_2O

Each reaction was overlaid with mineral oil and heated to 95°C for 2 minutes.

0.5µl of 5u/µl Taq DNA polymerase was added to each reaction, which was amplified for 25 repetitions of 95°C for 30 seconds, 65°C for 30 seconds, 72°C for 1 minute, followed by a final extension of 72°C for 5 minutes.

7.5µl of each reaction was subjected to electrophoresis on a 1.5% agarose gel stained with ethidium bromide. The negative control reactions that lacked DNA generated no product, while those reactions containing all components generated a smear of products of various molecular weights. In contrast, the reactions containing DNA but no internal primer were incapable of generating product. These results confirmed that adapters had been ligated successfully to genomic fragments and all 3' ends capable of extension in the presence of a DNA polymerase had been terminated. The preferred method was termination prior to ligation since (i) this guaranteed that all fragments successfully ligating were terminated and (ii) the opportunities for inter-fragment ligation were remote.

Amplification of 5' and 3' flanking sequences from terminated, adapter-ligated genomic fragments.

Amplification reactions were performed for each VNTR primer containing the following:

5µl	5µl	10x Taq PCR buffer
5µl	5µl	10x dNTPs
2µl	2µl	25pmol/µl 24mer
2µl	2µl	25pmol/µl (AC)11B or (CA)11D or (GT)11H or (TG)11V
2µl	0µl	fragmented, terminated, adapter-ligated genome (approx. 50ng/µl)
34µl	36µl	dH_2O
50µl	50µl	

In addition, a parallel reaction was prepared containing all components except a VNTR primer.

All reactions were overlaid with mineral oil and heated to 95°C for 2 minutes. 0.5µl of 5u/µl Taq DNA polymerase was added to each tube and amplification was achieved by thermal cycling for 18 repetitions of 95°C for 30 seconds, 65°C for 45 seconds, 72°C for 45 seconds, followed by a final extension of 5 minutes at 72°C.

5µl of each reaction was loaded onto a 1.5% agarose gel stained with ethidium bromide, along with a molecular weight marker. The reactions that contained all components generated a smear of products of ranging from approximately 100 to 500bp, the intensity and distribution of molecular weights being comparable for each reaction. The lanes corresponding to those reactions lacking DNA and the reaction lacking a VNTR primer did not contain any product of amplification.

Example 3

The efficiency of digestion of the repeat sequence from a VNTR primed PCR product by T4 DNA polymerase was assessed.

A cloned VNTR allele was amplified by Taq DNA polymerase and separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between

episodes of centrifugation. A volume of 40µl was recovered, the concentration of which was judged by agarose gel electrophoresis to be 130ng/µl, approximating to 1.3pmol/µl.

A 1.5u/µl dilution of T4 DNA polymerase was prepared with dH_2O. The amplified DNA was digested at a concentration of 0.3pmol/µl with varying concentrations of T4 DNA polymerase at 12°C:

1.5µl	10x T4 DNA polymerase buffer
0.75µl	10mM dATP
0.75µl	10mM dCTP
3.5µl	DNA
0, 0.5, 1, 2, or 4µl	1.5u/µl T4 DNA polymerase
to 15µl	dH_2O

Parallel reactions were prepared that lacked dNTPs. The reactions were incubated at 12°C for 1 hour, followed by heat inactivation at 70°C for 20 minutes.

7.5µl of each reaction were subjected to electrophoresis on a 2.5% agarose gel stained with ethidium bromide. In the absence of dNTPs all DNA was digested with enzyme concentrations exceeding 0.05u/µl. By contrast, there was no discernible loss of DNA in the presence of dNTPs at any concentration of T4 DNA polymerase.

The efficiency of digestion of the repeat sequence from a VNTR primed PCR product by T7 gene 6 exonuclease was assessed.

A cloned VNTR allele was amplified with the plasmid specific sense primer and the (GT)11H primer by Taq DNA polymerase in the presence of [α-33P] dATP. Parallel reactions were performed for primers that contained or lacked a succession of four phosphorothioate bonds. In the primer pair containing phosphorothioate bonds these where located at the 5' end of the plasmid specific primer and at the 3' end of the (GT)11H primer.

The amplified DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. Equal amounts of the amplification reactions were digested by T7 gene 6 exonuclease at 37°C for 15 and 30 minutes, the concentration of DNA approximating to 0.1pmol/ul:

3.6µl	DNA
2µl	5x T7 gene 6 exonuclease buffer
1µl	10u/µl T7 gene 6 exonuclease
3.4µl	dH_2O
10µl	

A control reaction was incubated for 15 minutes at 37°C in the absence of enzyme.

All reactions were denatured at 95°C for 2 minutes with addition of 5µl formamide loading dye. 10µl of each sample was subjected to electrophoresis on an 8% polyacrylamide denaturing gel. An autoradiography film (Biomax MR; Kodak) was exposed to the gel after it had been fixed and dried.

It was found that after 15 minutes of incubation the DNA that lacked phosphorothioate protection had been digested completely. By contrast, the presence of phosphorothioate bonds preserved the DNA, one strand in each molecule becoming shortened by digestion of the enzyme, although some non-specific loss of DNA was seen.

The efficiency and specificity of digestion by T4 endonuclease VII and S1 nuclease was compared.

Cloned VNTR alleles of the same VNTR that differed in their repeat lengths by 4 nucleotides were amplified separately in the presence of [α-33P] dATP. The products derived from the shorter allele were divided equally between two tubes. To one tube an equal amount of the longer

allele was added and the mixture was hybridised by denaturing at 98°C for 2 minutes and annealing at 75°C for 150 minutes in 100mM NaCl and 200μM CTAB.

The hybridised and non-hybridised pools of DNA were separated from other low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation.

T4 endonuclease VII was diluted to 250u/μl in the supplied dilution buffer. Dilutions of S1 nuclease were prepared in dH_2O. Equal amounts of either hybridised DNA or non-hybridised DNA were digested by 50u/μl T4 endonuclease VII in Taq DNA polymerase buffer or by various concentrations of S1 nuclease in the supplied buffer. The S1 nuclease was added to the reactions to give final concentrations of 0.01u/μl, 0.03u/μl, 0.1u/μl, and 0.3u/μl. In each case a control reaction that lacked enzyme was prepared. The reactions were performed at 37°C for 30 minutes.

On completion of digestion the reactions were stopped by addition of EDTA and heat inactivation. An amount of formamide loading dye equal to half the reaction volume was added and each reaction was denatured by incubation at 95°C for 5 minutes. 12 μl of each sample were subjected to electrophoresis on an 8% polyacrylamide denaturing gel. An autoradiography film (Biomax MR; Kodak) was exposed to the fixed and dried gel.

T4 endonuclease VII was found to cleave about half of all DNA derived from hybridisation of approximately equal amounts of two different alleles of the same VNTR, creating a characteristic pattern of cleaved products corresponding to the position of the mis-match within the repeat sequence at the time of cleavage. The DNA derived from the single allele that had not been hybridised and, therefore, comprised mis-match free double stranded DNA was not affected by T4 endonuclease VII. In contrast, the characteristic pattern of cleaved products that was seen with T4 endonuclease VII was not seen in association with S1 nuclease under

any of the reaction conditions. As such, T4 endonuclease VII was considered the better of the two enzymes in this application.

Repetition of the T4 endonuclease VII reactions using various concentrations of enzyme for 30 minutes and 1 hour of digestion in 1x Taq PCR buffer, 1x Pfu buffer (Stratagene) and 1x T7 gene 6 exonuclease buffer confirmed that the enzyme digested predictably and reproducibly over a range of reaction conditions, their being no overt non specific digestion of DNA detectable at concentrations up to 200u/μl. The enzyme was found to cleave hybridised molecules containing mis-matches of a range of sizes.

The characteristic pattern of cleaved products resulting from a mis-match within a repeat sequence was seen with S1 nuclease only when large amounts of DNA were loaded onto a polyacrylamide gel. This was seen with a four nucleotide mis-match. The ability of S1 nuclease to resolve a two nucleotide mis-match was found to be poor.

The effect of enzyme concentration on the efficiency of cleavage of mis-match containing duplex DNA by T4 endonuclease VII was assessed.

Two cloned VNTR alleles that differed in allele length by 2 nucleotides were amplified separately using the plasmid specific primers, one of which had been labelled with [γ-33P] ATP using T4 polynucleotide kinase. Each amplified allele was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation.

Half of the DNA derived from amplification of the smaller allele was saved. To the remaining half was added approximately an equal amount of amplified DNA of the larger allele. This mixture was denatured at 98°C for 2 minutes and then annealed at 75°C for 2 hours in the presence of 100mM NaCl and 200μM CTAB, the transition between temperatures occurring rapidly. Separation of the annealed DNA from low

molecular weight solutes by microconcentration was repeated.

Serial dilutions of T4 endonuclease VII were prepared in the supplied dilution buffer. The non-denatured smaller allele and the allele mixture that had been denatured and annealed were each digested in Taq DNA polymerase buffer with T4 endonuclease VII at final concentrations of 0u/μl, 50u/μl, 100u/μl and 150u/μl:

6μl	DNA
1μl	10x Taq PCR buffer
3μl	T4 endonuclease VII
10μl	

Incubation at 37°C was carried out for 30 minutes, after which each reaction was heated to 95°C for 2 minutes with addition of 5μl formamide loading dye. 10μl volumes were subjected to electrophoresis on an 8% polyacrylamide denaturing gel, after which the gel was fixed, dried and exposed to an autoradiography film (Biomax MR; Kodak).

Almost no digestion of the non-denatured smaller allele was detected. The little that was seen was assumed to have occurred as a result of digestion at sites of polymerase error or the annealing of stutter bands during the final cycle of amplification. In the lanes corresponding to the annealed allele mixture the characteristic pattern of digestion was seen to occur in the presence of T4 endonuclease VII. Although the amount of digestion at 100u/μl appeared to be slightly greater than at 50u/μl, the degree of digestion at each enzyme concentration was found to be almost uniform.

Similar experiments were performed using various concentrations of T4 endonuclease VII in Pfu buffer (Stratagene) and T7 gene 6 exonuclease buffer. Efficient digestion of mis-match containing DNA was found to occur in both reaction buffers, the degree of digestion maximising at concentrations of T4 endonuclease VII between 50u/μl and 100u/μl. Duplex DNA lacking a mis-match was resistant to T4

endonuclease VII under these conditions.

The efficiency and specificity of S1 nuclease digestion in T7 gene 6 exonuclease buffer was assessed.

A cloned VNTR allele was amplified with the plasmid specific primers, one of which had been labelled with [γ-33P] ATP using T4 polynucleotide kinase. The amplified product was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. The volume of recovered DNA was divided: 30μl was preserved as double stranded DNA while the remaining 30μl DNA was rendered single stranded by denaturation at 98°C for 2 minutes followed by snap cooling on iced water.

Dilutions of S1 nuclease were prepared in dH_2O. Equal amounts of double stranded DNA or single stranded DNA were digested in T7 gene 6 exonuclease buffer at 37°C for 5 minutes in the presence of S1 nuclease at final concentrations of 0u/μl, 0.1u/μl, 0.3u/μl, 1u/μl and 3u/μl. On completion of digestion the reactions were stopped by addition of 500mM EDTA pH8 to a final concentration of 25mM.

The reactions were denatured by addition of formamide loading dye and heating to 95°C for 3 minutes, after which aliquots were subjected to electrophoresis on an 8% polyacrylamide denaturing gel. The gel was fixed, dried, and exposed to an autoradiography film (Biomax MR; Kodak).

It was found that a concentration of 1u/μl S1 nuclease in T7 gene 6 exonuclease buffer produced optimal digestion of single stranded DNA, there being no overt loss of double stranded DNA at this concentration.

Assessment of the digestion of DNA by T7 gene 6 exonuclease in concert with S1 nuclease.

For assessment of T7 gene 6 exonuclease and S1 nuclease, DNA was amplified from a cloned VNTR allele using the plasmid specific sense primer with four phosphorothioate bonds at the 5' end and either the (AC)11B primer containing four phosphorothioate bonds at the 3' end or the (AC)11B primer that lacked such bonds. The amplified products were separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. The volumes recovered in each case were measured to be 40µl. These were found to contain approximately 1.3pmol/µl and 0.35pmol/µl for the reactions primed by the VNTR primer with and without phosphorothioate bonds, respectively.

T7 gene 6 exonuclease was diluted to 10u/µl in dH_2O.

S1 nuclease was diluted to 10u/µl in dH_2O.

Each amplified product, at a concentration of approximately 0.1pmol/µl, was digested by T7 gene 6 exonuclease. In addition, the DNA generated with the (AC)11B primer containing phosphorothioate bonds was digested by T7 gene 6 exonuclease in concert with S1 nuclease:

without PT bonds		with PT bonds	with PT bonds	
4µl	4µl	4µl	5x T7 gene 6 buffer	
5.7µl	1.6µl	1.6µl	DNA	
0, 2, 4, 8µl		0, 2, 4, 8µl	0, 2, 4, 8µl	10u/µl T7 gene 6 exonuclease
0µl	0µl	2µl	10u/µl S1 nuclease	
to 20µl	to 20µl	to 20µl	dH_2O	

Each reaction was incubated at 37°C for 10 minutes, after which 1µl 500mM EDTA pH8 was added to each tube followed by incubation at 70°C for 20 minutes.

10µl of each digest was subjected to electrophoresis on a 2.5% agarose gel stained with ethidium bromide. Lanes corresponding to reactions lacking enzyme contained a discrete band of the expected

molecular weight. The appearance of a lower molecular weight band, corresponding to single stranded DNA, was seen at a concentration of 1u/μl T7 gene 6 exonuclease for DNA primed by the (AC)11B primer that lacked phosphorothioate protection. At concentrations exceeding this virtually all DNA was single stranded. In contrast, DNA protected by phosphorothioate bonds at each end did not appear to alter significantly in molecular weight at any of the concentrations of T7 gene 6 exonuclease, but a decrease in the amount of DNA was evident with increasing concentrations. Similarly, DNA protected at each end was resistant to digestion of T7 gene 6 exonuclease in combination with S1 nuclease. Concentrations of 1u/μl T7 gene 6 exonuclease with 1u/μl S1 nuclease in T7 gene 6 exonuclease buffer containing approximately 0.1 pmol/μl DNA appeared to give the best results.

The mis-match discrimination procedure was assessed using a model system comprising three alleles of the same VNTR in concert with a single allele of a second VNTR.

A mixture of VNTR alleles was prepared that contained three alleles of the same VNTR, (AC)10, (AC)11, and (AC)18, in a 2 : 1 : 1 ratio respectively. In addition, an amount of the (CA)16 allele of a second VNTR, equal to that of the (AC)11 and (AC)18 alleles, was added to the mixture. Using Pfu DNA polymerase (Stratagene) 1ng of the mixture was amplified by PCR in a reaction volume of 100μl containing 60 pmoles of each plasmid specific primer, the sense primer having been labelled with [γ-33P] ATP. Thermal cycling was performed for 17 repetitions of 95°C for 30s, 65°C for 30s, 72°C for 45s, followed by a final extension of 72°C for 5 minutes.

The amplified DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation. The recovered DNA was denatured at 98°C for 2 minutes and then annealed at 75°C for 2 hours in 100mM NaCl

and 200μM CTAB, the transition between temperatures being rapid.

The hybridised DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation, and digested by T4 endonuclease VII in Taq DNA polymerase buffer containing 50u/μl of the enzyme in a total volume of 36μl. Digestion proceeded at 37°C for 1 hour after which the reaction was incubated at 75°C for 15 minutes.

The digested DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation. Further digestion was performed in a 50μl reaction containing 1u/μl T7 gene 6 exonuclease and 1u/μl S1 nuclease in T7 gene 6 exonuclease buffer at 37°C for 10 minutes. The reaction was stopped by addition of 2μl 500mM EDTA pH8 and heating to 75°C for 10 minutes.

Microconcentration was performed (Microcon-30; Amicon) with addition of dH_20 between episodes of centrifugation. A volume of 48μl was recovered of which 4μl was amplified by PCR, as before. This was followed by a second round of the mis-match discrimination procedure.

Aliquots of the amplified DNA before and after each round of the mis-match discrimination procedure were subjected to electrophoresis on an 8% polyacrylamide denaturing gel. In addition, for comparison of the molecular weight of each product, the PCR products of each allele amplified in isolation were loaded onto the gel.

It was found that Pfu generated numerous stutter bands in each amplification reaction. The amount of the (AC)10 allele in the mixture prior to mis-match discrimination was approximately twice that of all other alleles. These others were present in approximately equal amounts. After the first round of mis-match discrimination obvious enrichment of the (AC)10 allele was seen. This was enhanced by the second round of mis-match discrimination giving rise to a very strong band corresponding to the

(AC)10 allele and marked reduction of the (AC)11 and (AC)18 alleles. Although a band corresponding to the (CA)15 allele of the second VNTR was present after the second round of mis-match discrimination it was not as bright as that of the enriched (AC)10 allele. This was considered to reflect the inequality in the total DNA of each VNTR within the mixture and the consequential relative inefficiency of hybridisation following second order kinetics. This experiment confirmed that mis-match discrimination enriches the allele in a mixture of alleles of the same VNTR that has the highest frequency.

Example 4

The protocol was assessed using the pooled genomes of several dogs.

In the absence of DNA samples from individuals affected and unaffected by a hereditary trait the protocol was validated on a model system designed to mimic a scenario of VNTR linkage disequilibrium that would be expected in the presence of a recessive trait.

A total of 43 dogs were genotyped with respect a VNTR previously isolated in the dog using VNTR specific primers. The VNTR primer pair comprised (CACTTGGGACTTTGGATTGGTCA)$_\lambda$ SEQ ID NO:7 sense primer and (GTCTTTGTTTCCATTCTTGCTTGC)$_\lambda$ SEQ ID NO:8 antisense primer.

Amplification reactions by PCR were performed in a volume of 10µl containing 20ng genomic DNA and 4pmoles of each VNTR specific primer. In each case the VNTR specific sense primer was labelled and added to an amplification reaction master mix:

1.5µl	10x T4 polynucleotide kinase buffer
2.4µl	50pmol/µl VNTR specific sense primer
4.5µl	[γ-33] ATP
1µl	1 in 3 dilution of 30 u/µl T4 polynucleotide kinase
5.6µl	dH_2O
15µl	

The reaction was incubated at 37°C for 1 hour, then 90°C for 5 minutes.

The T4 polynucleotide kinase reaction was added to a PCR master mix:

15µl	T4 polynucleotide kinase reaction
45µl	10x Taq DNA polymerase buffer
45µl	10x dNTPs
2.4µl	50pmol/µl VNTR specific antisense primer
4.5µl	5u/µl Taq DNA polymerase
293µl	dH_2O
405µl	

For each dog 1µl of 20ng/µl genomic DNA was added to 9µl of PCR master mix which was overlaid with mineral oil. Each reaction was placed onto a preheated thermal cycler at 95°C and incubated for 2 minutes. Thermal cycling then followed with 28 repetitions of denaturation at 95°C for 30s, annealing at 65°C for 30s, and extension at 72°C for 30s, followed by a final extension of 72°C for 5 minutes.

On completion of thermal cycling 5µl of formamide loading dye was added to each reaction with denaturation at 90°C for 3 minutes prior to electrophoresis at 60W on an 8% polyacrylamide denaturing gel. The gel was fixed in 10% methanol/10% glacial acetic acid and dried. An autoradiography film (BioMax MR; Kodak) was exposed to the gel overnight.

The genotype of each dog was scored with respect to the VNTR. Ten dogs were selected to represent the 'affected pool' of individuals and ten were selected to represent the 'wild type pool'. This selection was made in order to achieve a scenario that may mimic a recessive trait:

Affected	Allele frequency
(AC)n	100%
(AC)n+1	0%
(AC)n+2	0%
(AC)n+3	0%
(AC)n+4	0%
(AC)n+5	0%
(AC)n+6	0%
(AC)n+7	0%

Wild type	Allele frequency
(AC)n	15%
(AC)n+1	0%
(AC)n+2	0%
(AC)n+3	0%
(AC)n+4	35%
(AC)n+5	20%
(AC)n+6	0%
(AC)n+7	30%

Amplimers were prepared from genomic DNA of a single dog. In a 100µl volume 5µg of genomic DNA were digested by 20 units Hae III, the digestion proceeding to completion over 12 hours at 37°C:

4.4µl	1.135µg/µl genomic DNA
10µl	10x restriction buffer
2µl	10u/µl Hae III
84µl	dH_2O
100µl	

Digestion was confirmed by electrophoresis of an aliquot of the reaction on a 1% agarose gel stained with ethidium bromide.

The DNA was extracted (GFX purification column) and eluted

in 50µl 5mM Tris pH8.5, of which approximately 3µg contained within 30µl was incubated with Terminal deoxynucleotidyl transferase for 3 hours at 37°C :

30µl DNA
30µl 5x Terminal deoxynucleotidyl transferase buffer
4.5µl 10mM ddGTP
10µl 9u/µl Terminal deoxynucleotidyl transferase
75.5µl dH_2O
150µl

The DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. A volume of 35µl was recovered.

An adapter was prepared by annealing two oligonucleotides, a 24mer (GsCsAsGsGAGACATCGAAGGTATGAAC, where 's' represents SEQ ID NO: 4 a phosphorothioate bond) and a 12mer (TTCATACCTTCG) SEQ ID NO: 5

7.6µl 197pmol/µl 24mer
9.2µl 162pmol/µl 12mer
1.87µl 10x T4 DNA ligase buffer
18.7µl

The mixture was heated to 55°C and allowed to cool to 10°C over one hour.

The adapter was ligated to the terminated genomic fragments:

35µl DNA
18.7µl adapter
4.3µl 10x T4 DNA ligase buffer
1.5µl 10u/µl T4 DNA ligase
2.5µl dH_2O
62µl

The reaction was incubated at 16°C over night, then heat inactivated at 70°C for 20 minutes.

The DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. A volume of 54µl was recovered.

To prevent generation of spurious products through priming from sites of single strand nicks, these were terminated by incubation with Thermo Sequenase:

54µl	DNA
4.4µl	Thermo Sequenase buffer
1.4µl	10mM ddATP
1.4µl	10mM ddCTP
1.4µl	10mM ddGTP
1.4µl	10mM ddTTP
0.5µl	32u/µl Thermo Sequenase
5.5µl	dH_2O
70µl	

The mixture was overlaid with mineral oil and incubated at 74°C for 2 hours.

The DNA was extracted (GFX purification column) and eluted in 50µl 5mM Tris pH 8.5.

Amplimers were prepared from this DNA using VNTR primers and the 24mer oligonucleotide contained within the adapter as the adapter primer:

5µl	10x Taq DNA polymerase buffer
5µl	10x dNTPs
2µl	25pmol/µl adapter primer
2µl	25pmol/µl VNTR primer [(AC)11B, (CA)11D, (GT)11H, or (TG)11V]

2μl terminated, adapter-ligated DNA fragments (approx. 50ng/μl)

34μl dH_2O

50μl

Similar reactions were prepared containing a VNTR primer but in the absence of genomic DNA. In addition, a single reaction was performed containing genomic DNA but in the absence of a VNTR primer. All reactions were overlaid with mineral oil and incubated at 95°C for 2 minutes. Addition of 0.5μl of 5u/μl Taq DNA polymerase was made to each reaction. Amplification was achieved by thermal cycling for 18 repetitions of 95°C for 30 s, 65°C for 45s, 72°C for 45s, followed by a final extension of 72°C for 5 minutes.

On completion of amplification 5μl of each reaction were subjected to electrophoresis with a molecular weight marker on a 1.5% agarose gel stained with ethidium bromide. The presence of amplified products in the lanes representing reactions containing template DNA and a VNTR primer confirmed that ligation of the genomic fragments to adapter sequence had occurred. In each case the appearance of these lanes was similar, there being a smear of amplified products distributed over a range of molecular weights from approximately 100bp to 500bp. All other lanes lacked product of amplification. The fact that the reaction containing template DNA but no VNTR primer did not generate product confirmed that the all 3' ends had been terminated successfully such that chain extension in the presence of Taq DNA polymerase was prevented.

The (AC)11B and (CA)11D primed reactions were combined. Also, the (GT)11H and (TG)11V primed reactions were combined. Both amplimer pools were separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation. Quantification by agarose gel electrophoresis of the recovered DNA suggested that each contained approximately 35ng/μl

amplimer DNA.

The repeat sequences were removed from the pooled (AC)11B and (CA)11D primed products using T4 DNA polymerase and Exonuclease VII:

14µl	35ng/µl (AC)11B/(CA)11D primed amplimer DNA
2µl	10x T4 DNA polymerase buffer
1µl	10mM dATP
1µl	10mM dCTP
2µl	1 in 4 dilution of 4u/µl T4 DNA polymerase
20µl	

The reaction was incubated at 12°C for 1 hour then inactivated at 70°C for 20 minutes.

To the reaction was added 1µl of 10u/µl Exonuclease VII with incubation at 37°C for 30 minutes followed by 70°C for 20 minutes.

The designated affected and wild type DNA pools were prepared by combining equal amounts of genomic DNA, quantified by spectrophotometry, of the selected dogs. These were phenol/chloroform extracted and microconcentrated (Microcon; Amicon) with addition of dH_2O between episodes of centrifugation.

Each pool of genomic DNA was digested by Hae III, terminated using Terminal deoxynucleotidyl transferase, and ligated to the adapter in a manner similar to that previously described. Complete termination of all 3' ends was confirmed by PCR with the adapter primer. The DNA pools were quantified by agarose gel electrophoresis and were found to contain approximately equal concentrations.

In a minimal volume 2.5µl of the 35ng/µl (AC)/(CA) primed amplimer pool, digested with T4 DNA polymerase and Exonuclease VII, were hybridised in 0.6M NaCl to approximately 300ng of the 'affected' genomic DNA pool that had been fragmented, terminated, and ligated to the adapter. This was achieved by denaturing the mixture under mineral oil

at 98°C for 3 minutes, followed by a stepwise reduction in the temperature from 80°C to 70°C over ten hours and sustaining the final temperature for a further 10 hours. The wild type pool was hybridised in a similar manner in parallel.

To each hybridisation were added:

20µl 10x Taq DNA polymerase buffer

20µl 10x dNTPs

160µl dH_2O

200µl

In each case the total volume containing the hybridised DNA was divided between two reaction tubes. Under mineral oil each volume was heated to 75°C. 1µl of 5u/µl Taq DNA polymerase was added to each tube followed by incubation at 72°C for 10 minutes.

The reactions were denatured at 95°C for 3 minutes and 4µl of 25pmol/µl adapter primer were added. Amplification of the hybridised DNA was achieved by thermal cycling for 30 repetitions of 95°C for 30s, 65°C for 30s, 72°C for 90s, followed by a final extension of 72°C for 5 minutes.

The reactions containing affected DNA were pooled, as were the reactions containing wild type DNA, and 8µl of 10u/µl Exonuclease I were added to each 200µl volume of amplified DNA. The reactions were incubated at 37°C for 15 minutes.

For each reaction the DNA was separated from low molecular weight solutes (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation. In each case a volume of 10µl was recovered. The alleles contained within each sample were denatured and allowed to anneal by incubation under mineral oil at 98°C for 5 minutes followed by a rapid reduction in temperature to 75°C. At 75°C 2M NaCl and 10mM CTAB were added to give final concentrations of 50mM and 500µM, respectively. The hybridisation reactions were incubated at 75°C for a further 16 hours.

To each hybridisation reaction was added 150µl of 5mM Tris

pH 8.5. The diluted hybridisation reactions were then separated from low molecular weight solutes (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation. These were judged to contain approximately 10pmoles DNA. Digestion by T4 endonuclease VII at a concentration of 50u/µl in Taq DNA polymerase buffer was performed in a volume of 100µl. The digestion proceeded at 37°C for 30 minutes prior to incubation at 65°C for 15 minutes.

Each digest was separated from low molecular weight solutes (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation. The recovered volume in each case was divided between three tubes, each being digested either by 0.5u/µl Exonuclease I in 1x Taq DNA polymerase buffer, 1u/µl T7 gene 6 exonuclease followed after heat inactivation at 70°C for 10 minutes by 0.5u/µl Exonuclease I in 1x T7 gene 6 exonuclease buffer, or 1u/µl T7 gene 6 exonuclease together with 1u/µl S1 nuclease in 1x T7 gene 6 exonuclease buffer. The concentration of DNA in each reaction was approximately 0.1pmol/µl contained within a 30µl volume. The Exonuclease I reactions were performed at 37°C for 15 minutes prior to heat inactivation at 70°C for 10 minutes. The reactions containing T7 gene 6 exonuclease with or without S1 nuclease were performed at 37°C for 10 minutes. On completion of each regime of digestion the DNA was extracted (GFX purification column) and eluted in 50µl dH_2O.

Three quarters of each of the extracted DNA samples was amplified by PCR with Taq DNA polymerase

37.5µl	digested DNA
15µl	10x Taq DNA polymerase buffer
15µl	10x dNTPs
6µl	25pmol/µl adapter primer
76.5µl	dH_2O
150µl	

The reactions were divided into 75µl aliquots and overlaid with mineral oil to which were added 0.75µl of 5u/µl Taq DNA polymerase after incubation at 95°C for 2 minutes. Amplification was achieved by thermal cycling for 25 repetitions of 95°C for 30s, 65°C for 30s, 72°C for 90s, followed by a final extension of 72°C for 5 minutes.

To each 150µl of amplified DNA were added 6µl 10u/µl Exonuclease I. The reactions were incubated at 37°C for 15 minutes.

The DNA in each case was separated from low molecular weight solutes (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation. Repetition of hybridisation in 50mM NaCl and 500µM CTAB followed by each regime of digestion was repeated, followed by amplification of the resulting DNA by PCR with Taq DNA polymerase, as above.

Aliquots of each of the amplified samples were subjected to electrophoresis on a 1.5% agarose gel stained with ethidium bromide with a molecular weight marker. The amplified products in the lanes corresponding to DNA digested by T4 endonuclease VII followed by Exonuclease I were of high molecular weight smearing towards the well. In contrast, the lanes corresponding to amplified product that had been digested by either T7 gene 6 exonuclease followed by Exonuclease I or T7 gene 6 exonuclease concomitantly with S1 nuclease contained products ranging in molecular weights from approximately 200bp to 750bp. The distribution of molecular weights in each case was similar. No smearing towards the well was seen suggesting that the spurious products of amplification that were seen in the absence of T7 gene 6 exonuclease were eliminated by the presence of this enzyme. As such, T7 gene 6 exonuclease was considered an essential component of the mis-match discrimination regime for removal of repeat sequences from T4 endonuclease VII cleaved molecules that would otherwise cross-hybridise and produce spurious DNA molecules.

To each of the 150µl volumes of amplified DNA resulting from

the second round of mis-match discrimination were add 6μl of 10u/μl Exonuclease I and the reactions were digested at 37°C for 15 minutes.

The DNA in each case was separated from low molecular weight solutes (Microcon-30; Amicon) with addition of dH_2O between episodes of centrifugation.

For each of the reactions corresponding to the 'affected' dogs amplification was performed by PCR with Taq DNA polymerase using the VNTR specific primers in a volume of 50μl containing approximately 25ng DNA. Amplification by 28 repetitions of thermal cycling was performed after which 5μl aliquots and a molecular weight marker were loaded onto a 2% agarose gel stained with ethidium bromide.

For the lanes corresponding to digestion by T4 endonuclease VII and Exonuclease I the product of the expected molecular weight was very faint. In addition a large amount of spurious product in the vicinity of the wells was seen. For all other lanes no high molecular weight products were seen. Furthermore, the amplified products were seen clearly as a discrete band of the expected molecular weights of approximately 130bp.

The products of amplification corresponding to digestion by T4 endonuclease VII and Exonuclease I were discarded. The remaining reactions were amplified further using the VNTR specific primers, one of which was labelled with [γ-33P] ATP using T4 polynucleotide kinase. Amplification reactions were performed by PCR using Taq DNA polymerase in volumes of 20μl containing 10pmoles of each primer for 35 repetition of thermal cycling. In addition, reactions were performed in the same manner containing 40ng of the pooled 'affected' and pooled 'wild type' DNA. After addition of 10μl of formamide loading dye to each sample the amplified products were denatured at 90°C for 3 minutes. 6μl aliquots of the mixture were subjected to electrophoresis on an 8% polyacrylamide denaturing gel. The gel was fixed and dried and exposed to an autoradiography film.

It was found that product was visible for DNA amplified from

affected DNA following the second round of mis-match discrimination. This was seen in both the lanes corresponding to digestion by T7 gene 6 exonuclease followed by Exonuclease I and those corresponding to digestion by T7 gene 6 exonuclease concomitantly with S1 nuclease. In each case the product resembled that resulting from amplification of the pooled affected DNA that had not been subjected to mis-match cleavage. In the case of wild type DNA amplified after the second round of mis-match discrimination no products were discernible.

This experiment confirmed that VNTRs are reproduced with fidelity from the pooled genomes of several individuals, the alleles in each case being preserved, and mis-match discrimination serves to eliminate spurious products of amplification and enrich the VNTR allele of the highest frequency. Although no products were visible for DNA derived from the wild type DNA, it may be that products would become visible with higher loading of DNA on the polyacrylamide gel. As such, further repetition of the mis-match discrimination procedure would be necessary to reduce to near homozygosity the alleles in both DNA pools such that final selection of the informative allele could be achieved.

Example 5

Demonstration of the resistance to Exonuclease III of DNA with a 3' overhang derived by ligation to an adapter.

A cloned VNTR allele was amplified by Taq DNA polymerase. The amplified DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation.

The volume recovered was measured at 44μl, the concentration of which was determined by agarose gel electrophoresis to be 160ng/μl, approximating to 1.6pmol/μl.

The amplified DNA was blunted by T4 DNA polymerase digestion:

42µl DNA
3.25µl 10mM dATP
3.25µl 10mM dCTP
3.25µl 10mM dGTP
3.25µl 10mM dTTP
13µl 10x T4 DNA polymerase buffer
3.25µl 4u/µl T4 DNA polymerase
59µl dH_2O
130µl

The reaction was incubated at 12°C for 30 minutes, then heat inactivated at 70°C for 20 minutes. The DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. A volume of 30µl was recovered.

1600pmoles of a 21mer oligonucleotide (CTCGCAAGGATGGGATGCTCG) SEQ ID NO:9 were phosphorylated with T4 polynucleotide kinase diluted to 10u/µl in the supplied dilution buffer:

3.19µl 21mer oligonucleotide
1.5µl 10x T4 DNA ligase buffer
1µl 10u/µl T4 polynucleotide kinase
9.3µl dH_2O
15µl

The reaction was incubated at 37°C for 30 minutes, then heat inactivated at 90°C for 10 minutes.

To the kinase reaction was added 1600pmoles of a 12mer oligonucleotide (CATCCTTGCGAG) SEQ ID NO:15 Annealing of the oligos to form an adapter was achieved by heating to 55°C and allowing the mixture to cool to 10°C over a period of 1 hour.

Half of the DNA blunted by T4 DNA polymerase was saved.

To the annealed adapter was added the remaining 15µl of blunted DNA

such that the adapter was in a 50 fold excess:

15µl blunted DNA

16.2µl annealed adapter

1.9µl 10x T4 DNA ligase buffer

1µl 10u/µl T4 DNA ligase

34µl

The ligation reaction was incubated over night at 16°C.

The ligation was heat inactivated at 70°C for 20 minutes and the DNA was separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation.

The volume recovered was measured to be 36µl.

The ligated DNA and 15µl of non-ligated DNA that had been saved were both made to approximately 0.75pmoles/µl by addition of dH_2O. Each was digested by Exonuclease III at a final concentration of DNA approximating to 0.2pmol/µl:

10.7µl DNA

4µl 10x Exonuclease III buffer

1µl 200u/µl Exonuclease III

24.3µl dH_2O

40µl

The reaction was incubated 37°C for 5 minutes then heat inactivated at 70°C for 20 minutes.

Approximately 2pmoles of each digest were loaded onto a 2% agarose gel stained with ethidium bromide. All non-ligated DNA was digested to completion by Exonuclease III such that none was detectable on the agarose gel. In contrast, although some digestion had occurred, much of the ligated DNA was found to be resistant to digestion. That which had been digested was assumed to have failed to ligate to the phosphorylated adapter. This experiment confirmed that ligation of an

adapter is one method by which DNA molecules may become resistant to Exonuclease III digestion, those molecules lacking an adapter being digested to completion by this enzyme.

Selection of unique sequences in a pool of DNA hybridised to a second pool of DNA using Exonuclease III.

Two cloned VNTR alleles that differed in their repeat lengths by four nucleotides were amplified by PCR using Taq DNA polymerase. The amplified DNAs were separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation and the resulting concentrations of DNA were determined by agarose gel electrophoresis.

To a portion of the amplified products of the smaller allele was added a 3' overhang by incubation with Terminal deoxynucleotidyl transferase:

12.5µl 120ng/µl DNA (approx. 1.2pmol/µl)
15µl 5x Terminal deoxynucleotidyl transferase buffer
1.125µl 10mM dATP
3.3µl 9u/µl Terminal deoxynucleotidyl transferase
43µl dH_2O
75µl

The reaction was incubated at 37°C for 1 hour after which the DNA was extracted (GFX purification column).

To 450ng of the allele possessing a 3' overhang was added:

(i) 4.5µg of the same allele that lacked a 3' overhang;

(ii) 4.5µg of the larger allele that lacked a 3' overhang.

In each case, the total volume was minimised by microconcentration (Microcon-30; Amicon). These mixtures were denatured at 98°C for 3 minutes and annealed at 75°C for 2 hours in the presence of 0.2M NaCl and 100µM CTAB.

To each hybridisation reaction were added:

10µl 10x Taq DNA polymerase buffer

10µl 500u/µl T4 endonuclease VII

80µl dH_2O

100µl

The reactions were incubate at 37°C for 45 minutes, then inactivated at 70°C for 15 minutes.

The DNAs were separated from low molecular weight solutes by microconcentration (Microcon-30; Amicon) with successive additions of dH_2O between episodes of centrifugation. In each case a volume of approximately 40µl was recovered which was diluted in a reaction mixture containing 5u/µl Exonuclease III:

40µl DNA

15µl 10x Exonuclease III buffer

3.75µl 200u/µl Exonuclease III

91µl dH_2O

150µl

The reactions were incubated at 37°C for 5 minutes, after which they were microconcentrated (Microcon-30; Amicon). The entire recovered volumes were subjected to electrophoresis on a 1.5% agarose gel stained with ethidium bromide. In addition, a molecular weight marker, 400ng of the small allele without a 3' overhang, and 400ng of the smaller allele that possessed an overhang were loaded on to the gel.

The size of the smaller amplified allele was confirmed to be approximately 150bp by comparison to the molecular weight marker. After incubation with Terminal deoxynucleotidyl transferase the apparent size of this amplified allele had increased. A smear of products distributed over a range of sizes corresponding to between 400bp and 750bp of double stranded DNA was seen, though the majority of DNA was confined to an ill-defined band midway between these. In the lane containing hybridised

alleles of different sizes that had been digested, a band corresponding to approximately 300bp of double stranded DNA was seen against a back ground smear of products. This band was considered to be the result of enzymatic cleavage of the mis-match containing DNA duplexes, where as the back ground smear was considered to be single stranded DNA resulting from Exonuclease III digestion of molecules lacking the protection of a 3' overhang. In the lane that contained hybridised alleles of the same size two ill-defined bands were visible against a background smear of products. The brightest band was of an appearance similar to that of the smaller allele following its incubation with Terminal deoxynucleotidyl transferase and was considered to represent the remaining single stranded DNA from heteroduplex molecules digested by Exonuclease III. The fainter band was considered to the result of enzymatic cleavage of molecules possessing polymerase errors. As before, the background smear was considered to be due to single stranded DNA of molecules lacking a 3' overhang that had resulted from digestion by Exonuclease III. This experiment suggests that an allele possessing a 3' overhang entering into a heteroduplex with an allele of a different repeat length is digested by T4 endonuclease VII and Exonuclease III such that a fragment of the heteroduplex may be selected.

Appendix

Consider a scenario that may typify a rare recessive trait. The affected group of individuals are homozygous for the same allele. In the wild type group, this allele has a relatively low frequency.

Starting scenario		**Affected**				**Wild Type**		
Alleles	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Allele frequencies	1.0	0.0	0.0	0.0	0.15	0.35	0.2	0.3
Allele ratios	1	0	0	0	3	7	4	6

After 1st Round		**Affected**				**Wild Type**		
Alleles	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Amount remaining	1.000	0.000	0.000	0.000	0.023	0.123	0.040	0.090
Total remaining	1.0				0.276			
Allele ratios	1	0	0	0	23	123	40	90
Allele frequencies	1.000	0.000	0.000	0.000	0.083	0.446	0.145	0.326

After 2nd Round		**Affected**				**Wild Type**		
Alleles	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Amount remaining	1.0	0.0	0.0	0.0	0.006	0.199	0.021	0.106
Total remaining	1.0				0.332			
Allele ratios	1	0	0	0	6	199	21	106
Allele frequencies	1.0	0.0	0.0	0.0	0.018	0.599	0.063	0.319

After 3rd Round		**Affected**				**Wild Type**		
Alleles	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Amount remaining	1.0	0.0	0.0	0.0	0.000	0.359	0.004	0.102
Total remaining	1.0				0.465			
Allele ratios	1	0	0	0	0	359	4	102
Allele frequencies	1.0	0.0	0.0	0.0	0.000	0.772	0.008	0.219

After 4th Round	Affected				Wild Type			
Alleles	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Amount remaining	1.0	0.0	0.0	0.0	0.000	0.596	0.000	0.010
Total remaining	1.0				0.606			
Allele ratios	1	0	0	0	0	596	0	10
Allele frequencies	1.0	0.0	0.0	0.0	0.000	0.983	0.000	0.017

Comparison of the ratios of remaining alleles

1 x 1 x 1 x 1= 1 0.276 x 0.332 x 0.465 x 0.606 = 0.026

38.5 : 1

all of which is A

none of which is A

Therefore, even if an large excess of wild type DNA is hybridised to the affected DNA that survives the mis-match discrimination procedure it is extremely likely that the allele present in the affected group will be recovered.

Consider another scenario in which one allele is present in the affected group of individuals at a frequency greater than that of the wild type group.

Starting scenario	Affected					Wild Type				
Alleles	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Allele frequencies	0.050	0.100	0.000	0.150	0.700	0.250	0.200	0.150	0.250	0.150
Allele ratios	1	2	0	3	14	5	4	3	5	3

After 1st Round	Affected					Wild Type				
Alleles	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Amount remaining	0.003	0.010	0.000	0.023	0.490	0.063	0.040	0.023	0.063	0.023
Total remaining	0.526					0.212				
Allele ratios	3	10	0	23	490	63	40	23	63	23
Allele frequencies	0.006	0.019	0.000	0.044	0.932	0.297	0.189	0.108	0.297	0.108

After 2nd Round	Affected					Wild Type				
Alleles	A	B	C	D	E	A	B	C	D	E
Amount remaining	0.000	0.000	0.000	0.002	0.869	0.088	0.036	0.012	0.088	0.012
Total remaining	0.871					0.236				
Allele ratios	0	0	0	2	869	22	9	3	22	3
Allele frequencies	0.000	0.000	0.000	0.002	0.998	0.373	0.153	0.051	0.373	0.051

After 3rd Round	Affected					Wild Type				
Alleles	A	B	C	D	E	A	B	C	D	E
Amount remaining	0.000	0.000	0.000	0.000	0.996	0.139	0.023	0.003	0.139	0.003
Total remaining	0.996					0.307				
Allele ratios	0	0	0	0	1	139	23	3	139	3
Allele frequencies	0.000	0.000	0.000	0.000	1.000	0.453	0.075	0.010	0.453	0.010

After 4th Round	Affected					Wild Type				
Alleles	A	B	C	D	E	A	B	C	D	E
Amount remaining	0.000	0.000	0.000	0.000	1.000	0.205	0.006	0.000	0.205	0.000
Total remaining	1.0					0.416				
Allele ratios	0	0	0	0	1	205	6	0	205	0
Allele frequencies	0.000	0.000	0.000	0.000	1.000	0.493	0.014	0.000	0.493	0.000

Comparison of the ratios of remaining alleles	0.526 x 0.871 x 0.996 x 1 = 0.456	0.212 x 0.236 x 0.307 x 0.416 = 0.006
	76 : 1	
	all of which is E	none of which is E

Therefore, even if an large excess of wild type DNA is hybridised to the affected DNA that survives the mis-match discrimination procedure it is extremely likely that allele E present in the affected group will be recovered.

References

Bruford M W, and Wayne R K (1993) Microsatellites and their application to population genetic studies. *Current Opinion in Genetics and Development.* **3**; 939-943.

Callen D F, Thompson A D, Phillips H A, Richards R I, Mulley J C, and Sutherland GR (1993) Incidence and origin of 'null' alleles in the (AC)n microsatellite markers. *Am J Hum Genet.* **52**; 922-927.

Murphy G (1993) Generation of a nested set of deletions using Exonuclease III. *Methods in Molecular Biology.* **23**; 51-59.

Clark D and Steven Henikoff (1994) Ordered deletions using Exonuclease III. *Methods in Molecular Biology.* **31**; 47-55.

Cooney A J (1997) Use of T4 DNA polymerase to create cohesive termini in PCR products for subcloning and site-directed mutagenesis. *BioTechniques.* **24**; 30-34.

Epplen J T, Buitkamp J, Bocker T and Epplen C (1995) Indirect gene diagnoses for complex (multifactorial) disease- a review. *Gene* **159**; 49-55.

Esteban J A, Salas M, and Blanco L (1992) Activation of S1 nuclease at neutral pH. *Nucleic Acids Research.* **20**; (18): 4932.

Hearne C M, Ghosh S, and Todd J A (1992) Microsatellites for linkage analysis of genetic traits. *Trends Genet.* **8**; (8): 288-294.

Karp A, Seberg O and Buiatti M (1996) Molecular Techniques in the Assessment of Botanical Diversity. *Annals of Botany* **78**; 143-149.

Lisitsyn N A (1995) Representational difference analysis: finding the differences between genomes. *Trends in Genetics*, **11**; 303-307.

Lu J, Knox M R, Ambrose M J and Brown J K M (1996) Comparative analysis of genetic diversity in pea assessed by RFLP- and PRC-based methods. *Theoretical and Applied Genetics* **93**; 1103-1111.

Mackill DJ, Zhang Z, Redona E D and Colowit P M (1996) Level of polymorphism and genetic mapping of AFLP markers in rice.

Genome **39**; 969-977.

Molyneux K, and Batt R M (1994) Five polymorphic canine microsatellites. *Animal Genetics*. **25**; 379.

Murphy G (1993) Generation of a nested set of deletions using Exonuclease III. *Methods in Molecular Biology*. **23**; 51-59.

Nelson SF, McCusker JH, Sander MA Kee Y, Modrich P, and Brown PO (1993) Genomic mis-match scanning: a new approach to genetic linkage mapping. *Nature Genetics* **4**; 11-18.

Nikiforov T T, Rendle R B, Kotewicz M L, and Rogers Y (1994) Use of phosphorothioate primers and exonuclease hydrolysis for the preparation of single-stranded PCR products and their detection by solid phase hybridisation. *PCR Methods and Applications*. **3**; 285-291.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The Comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2**; 225-238.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*. **23**; 4407-4414.